

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-333956
 (43)Date of publication of application : 22.11.2002

G06F 3/06

(51)Int.CI.

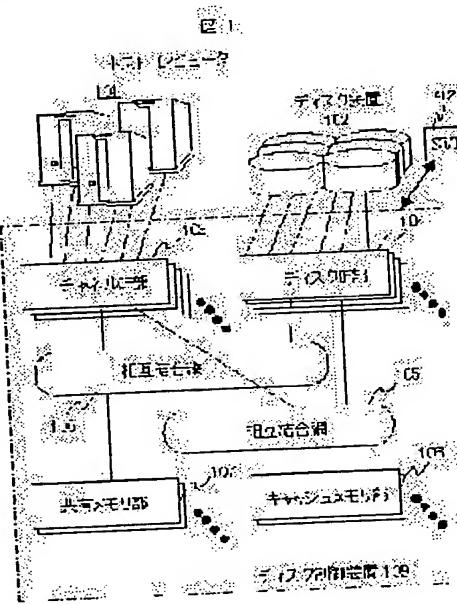
(21)Application number : 2001-138424
 (22)Date of filing : 09.05.2001

(71)Applicant : HITACHI LTD
 (72)Inventor : FUJIBAYASHI AKIRA
 FUJIMOTO KAZUHISA
 KANAI HIROKI

(54) COMPUTER SYSTEM USING DISK CONTROLLER AND ITS OPERATION SERVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To reduce the cost, and to effectively reflect the effects of the number of unit disk controllers in the performance when a plurality of unit disk controllers are allowed to function as one disk controller through a common interconnected network. SOLUTION: The access frequency of a logical volume is monitored, and a host device is encouraged to change an access path, and the move or copy of the logical volume is set in each unit disk controller so that an interconnected network can be used mainly for the copy of the logical volume, and that a necessary band can be suppressed low.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] Host computer. The channel-interface section which has one or two or more processors for controlling access with this host computer. Magnetic disk unit. The disk interface section which has one or two or more processors for controlling access with this magnetic disk unit. The function to connect mutually between the cache memory section, the aforementioned channel-interface section and the aforementioned disk interface section which store writing / data which carries out read-out between the aforementioned host computer and a magnetic disk unit, and the aforementioned cache memory sections. It is the computer system equipped with the above, and is characterized by having a means to total this access situation, and to make display an access frequency situation on a maintenance terminal as a means to supervise the access situation by the logical path to the aforementioned magnetic disk unit through the aforementioned common cross coupling network from the host computer of the aforementioned unit computer system, to a predetermined period or predetermined timing, or to notify a system administrator, in a computer system.

[Claim 2] Host computer. The channel-interface section which has one or two or more processors for controlling access with this host computer. Magnetic disk unit. The disk interface section which has one or two or more processors for controlling access with this magnetic disk unit. The function to connect mutually between the cache memory section, the aforementioned channel-interface section and the aforementioned disk interface section which store writing / data which carries out read-out between the aforementioned host computer and a magnetic disk unit, and the aforementioned cache memory sections. It is the computer system equipped with the above, and is characterized by having a means to total this access situation, and to make display an access frequency situation on a maintenance terminal as a means to supervise the access situation by the logical path to the aforementioned magnetic disk unit through the aforementioned common cross coupling network from the host computer of the aforementioned unit computer system, to a predetermined period or predetermined timing, or to notify a system administrator, in a computer system.

[Claim 3] Host computer. The channel-interface section which has one or two or more processors for controlling access with this host computer. Magnetic disk unit. The disk interface section which has one or two or more processors for controlling access with this magnetic disk unit. The function to connect mutually between the cache memory section, the aforementioned channel-interface section and the aforementioned disk interface section which store writing / data which carries out read-out between the aforementioned host computer and a magnetic disk unit, and the aforementioned cache memory sections. It is the computer system equipped with the above, and is characterized by having a means to total this access situation, and to make display an access frequency situation on a maintenance terminal as a means to supervise the access situation by the logical path to the aforementioned magnetic disk unit through the above 1st from the host computer of the aforementioned unit computer system, and the 2nd common cross coupling network, to a predetermined period or predetermined timing, or to notify a system administrator, in a computer system.

[Claim 4] The computer system according to claim 1 to 3 which has a means to replace with a

means to make display the access frequency situation of the logical volume of the magnetic disk unit through the aforementioned common cross coupling network, and a logical path on a maintenance terminal, or to notify a system administrator, and to report to high order equipment.

[Claim 5] The disk interface section characterized by providing the following. The cross coupling network which has the function to connect mutually between the cache memory section, the aforementioned channel-interface section and the aforementioned disk interface section which store writing / data which carries out read-out between the aforementioned host computer and a magnetic disk unit, and the aforementioned cache memory sections, It is applied to the computer system to which the unit computer system which makes a unit composition which becomes more was connected by the common cross coupling network connected through the aforementioned cross coupling network. While supervising the access situation by the logical path to the aforementioned magnetic disk unit through the aforementioned common cross coupling network from the host computer of the aforementioned unit computer system Employment service of the computer system characterized by totaling this access situation and proposing mapping change of a logical path and a disk side path to a system administrator or high order equipment according to an access frequency situation to a predetermined period or predetermined timing. Host computer. The channel-interface section which has one or two or more processors for controlling access with this host computer. Magnetic disk unit. One or two or more processors for controlling access with this magnetic disk unit.

[Claim 6] The disk interface section characterized by providing the following. The cross coupling network which has the function to connect mutually between the cache memory section, the aforementioned channel-interface section and the aforementioned disk interface section which store writing / data which carries out read-out between the aforementioned host computer and a magnetic disk unit, and the aforementioned cache memory sections, It is applied to the computer system to which the unit computer system which makes a unit composition which becomes more was connected by the common cross coupling network between meanses to connect the aforementioned host computer and the channel-interface section. While supervising the access situation by the logical path to the aforementioned magnetic disk unit through the aforementioned common cross coupling network from the host computer of the aforementioned unit computer system Employment service of the computer system characterized by totaling this access situation and proposing mapping change of a logical path and a disk side path to a system administrator or high order equipment according to an access frequency situation to a predetermined period or predetermined timing. Host computer. The channel-interface section which has one or two or more processors for controlling access with this host computer. Magnetic disk unit. One or two or more processors for controlling access with this magnetic disk unit.

[Claim 7] The disk interface section characterized by providing the following. The cross coupling network which has the function to connect mutually between the cache memory section, the aforementioned channel-interface section and the aforementioned disk interface section which store writing / data which carries out read-out between the aforementioned host computer and a magnetic disk unit, and the aforementioned cache memory sections, While the unit computer system which makes a unit composition which becomes more is connected by the 1st common cross coupling network connected through the aforementioned cross coupling network It is applied to the computer system connected by the 2nd common cross coupling network between meanses to connect the aforementioned host computer and the channel-interface section. While supervising the access situation by the logical path to the aforementioned magnetic disk unit through the aforementioned common cross coupling network from the host computer of the aforementioned unit computer system Employment service of the computer system characterized by totaling this access situation and proposing mapping change of a logical path and a disk side path to a system administrator or high order equipment according to an access frequency situation to a predetermined period or predetermined timing. Host computer. The channel-interface section which has one or two or more processors for controlling access with this host computer. Magnetic disk unit. One or two or more processors for controlling access with this magnetic disk unit.

with this magnetic disk unit.

[Claim 8] Employment service of a computer system according to claim 5 to 7 which replaces with mapping change of the aforementioned logical path and a disk side path, and proposes the copy and/or MUBU of an access place logical volume of a magnetic disk unit from a host computer.

[Claim 9] The logical path which performs automatically the copy and/or MUBU of change of the logical-path mapping concerned, and/or an access place logical volume in high order equipment according to mapping change of the logical path from the aforementioned disk controller, and a disk side path, the copy of an access place logical volume, and/or the contents of directions of MUBU corresponding to the aforementioned proposal, and a computer system [equipped with a means to manage correspondence of an access place logical volume] according to claim 5 to 8.

[Claim 10] The computer system according to claim 5 to 8 which displays and/or notifies simultaneously whether the procedure automatic directions means for performing the copy and/or MUBU of change of the logical-path mapping concerned and/or an access place logical volume with the aforementioned proposal according to mapping change of the logical path from the aforementioned disk controller and a disk side path, the copy of an access place logical volume, and/or the contents of directions of MUBU is started.

[Claim 11] The computer system according to claim 1 to 3 connected by the common cross coupling network which consists of two or more path cords which each between the aforementioned unit computer systems became independent of.

[Claim 12] It is the computer system according to claim 1 to 3 which boils the aforementioned common cross coupling network by the power supply which unified two power supplies which achieved [aforementioned] independence while driving by two power supplies which each of the aforementioned unit computer system became independent of, and is driven more.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any
error or omission in the text or drawings
caused by the use of this translation.

1. This document has been translated by computer. So the translation may not be precise.
2. *** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Section 1: Innovation]

[0001] [The technical field to which invention belongs] this invention relates to the computing system using the disk controller of the disk array which stores data in two or more disk units, and its

[0002] [Description of the Prior Art] When treating a lot of data now, various storage devices, such as a mainframe etc. by the data center etc., a tape backup unit, and the server group of an open system, are introduced and information is stored in each. Now, it is hard to tell each that those storing information is strongly in cooperation organizationally. Then, the concept of SAN (Storage Area Network) is introduced, various storage devices are connected in a network, and the movement referred to as managing information is becoming active. Although the large-scale data center was also building high reliance and the highly efficient and highly efficient storage system by the large-sized disk array until now, the disk array of a small medium size is also added and a larger-scale storage system than former is being built by SAN.

[0003] On the other hand, high performance and high reliability are required also in a small scale storage system on. Then, in the present disk array, while the disk array which is called

supports large-scale storage systems, such as small-scale and large-scale systems, which effectively is needed. It is needed the service which enables employment of this controller as shown in drawing 1. Two or more [0004]. The conventional disk array centers on a disk controller as shown in drawing 1. Two or more more disk interface which performs data transfer between a host computer 101, two or more channel-interface (following channel IF) sections 103 which perform data transfer between disk controllers 109, a magnetic disk unit 102, and a disk controller 109. The control information about the section 104, the cache memory section 107 which stores the data read and written in the section 104, the cache memory section 107, and the disk array control unit 109 ((IF) Following between Channel IF and the disk channel IF, and the disk array control unit 109 (which for example, stores the control information disk) It has the shared memory section 108 (which for example, stores the control information section 107 etc.). disk) It has the shared memory section 108 (which for example, stores the control information section 107 etc.). The shared memory section 108 is connected with the disk channel IF section 103 and the cache memory section 107 are about the data transfer of the channel IF section 103 and the cache memory section 107 are Each channel IF section 103, the disk IF section 104, and the cache memory section 107, the disk IF connected with the cross coupling network 105, and each channel IF section 103, the disk IF section 104, and the shared memory section 108 are also connected with the cross coupling section 104, and the shared memory section 108 are also connected with the cross coupling means, such as a network 106. The cross coupling network said here means all connecting means, and it collects switch, a loop, and a bus. Here, 412 is the SVP (maintenance terminal) section, and it collects the information on the access frequency for every channel path number of the information transmission between a host computer 101 and a disk controller 109 so that it may mention

needed at the time of extension is equipped, and thus, costs applying with basic composition. Equipment cost can be said to become comparatively high-priced inevitably at the basic composition time. Moreover, since the attachment for extended components of this ** also corresponds to improvement in the speed and the improvement in expandability (increase in the number of extensible components) in order to correspond to improvement in the speed of a host interface, and the improvement (increase in the connectable number of host interfaces) in KONEKUTIBITI, possibility of becoming high cost further and time of basic composition is large.

becoming comparatively high, process 100 is shown in [drawing 2](#), attaining the increase in efficiency of a system configuration according to the SAN (Storage Area Network) environment [0006]. On the other hand, as an outline is shown in [drawing 2](#), attaining the increase in efficiency of a system configuration according to the SAN (Storage Area Network) environment using the disk array is performed. The host computer group 101 is connected to a disk controller 109 through the common cross coupling network 210. Two or more channel IF sections 103, two or more disk IF sections 104, two or more shared memory sections 107, and two or more cache memory sections 108 have composition combined with the cross coupling networks 105 and 106, respectively, and the disk IF section has connected the disk controller 109 at it functions as a disk array. The disk unit 102 connected with a disk controller 109 at it functions as a disk array. Moreover, the common cross coupling network 210 can connect various storage devices, and the magnetic-tape-storage 212 grade is also connected. Specifically, all networks, such as a fiber channel switch, and a loop, LAN, can be considered. In the case of such a form (for example, small disk array equipment), it connects with dozens of sets and hundreds of sets, and it is possible as the aggregate of many logical volumes to show a system as the host computer of a high order, and high efficiency, such as a copy of the logical volume between disk controllers, large capacity equivalent to the large-sized disk array of high reliance and drawing. However, large disk array equipment has the problem that it is not what investigated high availability like large-sized disk array, and high-reliability. However, there is also a merit in the price side that is [drawing 1](#) in the SVP (maintenance terminal) sectu

is the aggregate of a cheap small disk array, ..., and collects the information on the access frequency for every channel path number of the information transmission between a host computer 101 and a disk controller 109. Furthermore, as analogous composition, as shown in drawing 3, it consists of cross coupling networks 105 and 106 which connect with the shared memory section 107 and the cache memory section 108 the channel IF section 103 which connects the host computer group 101 and the disk unit group 102, respectively, and the disk IF section 104. A disk controller shown in drawing 1 which functions as a disk array, and the disk controller shown in drawing 2 is used as the unit disk controller 309. This unit disk controller can be connected with a scale smaller than drawing 1, and the disk array which two or more sets and the common cross coupling network 310, and the disk array which functions as a disk controller 315 as a whole can also be constituted. In this case, a unit disk controller is considering as 1 / about four to 1/2 scale of a conventional-type disk controller, and mounting is also with a bird clapper compactly and can expect a cost fall. Moreover, the cost of the whole equipment can be fallen by preparing a necessary minimum band also for the common cross coupling network 310 which takes the lead. Here, 412 is the SVP (maintenance terminal) section and is a host computer 101, a disk controller 315, and a thing that collects the information for every channel path number of the information

information on the access to the unit disk controllers 309 more strictly. Furthermore, although illustration was omitted, there may also be a system which unified the system configuration shown in drawing 2 and drawing 3. That is, it joins together with the common cross coupling network 210 which shows between the host computers 101 in drawing 3 to drawing 2. If it does so, a host computer can access the unit disk controller linking directly to drawing 2. If the unit disk controller linking directly to self, other host computers, without minding the unit disk controller linking directly to the common cross coupling network via the cross coupling network between unit disk

Consequently, since direct access to controllers, access is improvable.

2003/05/10

https://www4.ipdlipo.go.jp/cgi-bin/tran_web.cgi_ejje

2003/08/10

the scalability to large-scale composition can be realized, it becomes unnecessary to constitute the mechanism at the time of extension in equipment beforehand, and an initial cost is reduced. However, if it sees by the relation between a host computer and a disk, the path about the cross coupling network between control units may incline, and this may check efficient employment of a system. Of course, although it becomes one solution to take a large band required for the cross coupling network between unit disk controllers, the merit of reduction of the initial cost by having used to prepare a big band in preparation for the bias of a specific path as the system which connects a unit disk controller with a cross coupling network is made reduced.

[0010] Furthermore, although it is necessary to take into consideration the path about the cross coupling network between control units, and redundantization of a power supply, this will also serve as an increase in cost simply, if the redundant path and redundant power supply for the cross coupling networks between control units are prepared simply.

[0011] It is in the purpose of this invention offering the computer system which enabled reduction of the data transfer between unit disk controllers as possible, and its employment service so that a band required for the cross coupling network between unit disk controllers in disk controller with the above-mentioned structure can be reduced, and if still more nearly required, suitable redundantization of the source of an electric power supply to this cross coupling network is proposed.

[0012] [Means for Solving the Problem] In order to attain the above-mentioned purpose and to improve the access path which a host computer uses, and the probability that the access place volume exists in the same unit disk controller, in this invention, As opposed to the high order equipment which supervises an access situation and connects it using the extracted access information which recommends or system administrator is displayed or notified of the information which has occurred recommends performing MUBL or the copy of the logical volume concerned which has occurred the passage between unit disk controllers frequently so that the optimal path may be used through a maintenance terminal, the web terminal for managers, etc. Moreover, it also makes it possible to perform them automatically. Thus, an exchange of the data between unit disk controllers is controlling the whole equipment so that it may mainly be concerned with the copy controllers is controlling a cross coupling network, and it stops a band required for a cross coupling network, and MUBL of volume, and it stops a band required for a cross coupling network being required also for redundantizing [0013] Moreover, from a common cross coupling network, only when two or more sets of controllers exist, using the power supply of two or more unit disk controllers, it is of the electric power supply to a common cross coupling network, only when two or more unit disk controllers exist, using the power supply of two or more unit disk controllers, it is considering as a redundant power supply, and the increase in an unnecessary power supply is

2003/06/10

<http://www4.ipdljpo.go.jp/cgi-bin/tranl?cmd=go>

2003/06/10

(maintenance terminal) section 412 and a system operator 411 a and the directions input means 410, and while being able to give a system operator 411 a notice/display for the signal of the output section 407 of the SVP (maintenance terminal) section 412, the directions from a system operator 411 are receivable in the input section 408. The configuration control section 406 receives the configuration information from a disk array, or directs a configuration change to a disk array. Moreover, it has a means 415 to transmit information mutually so that transmission and reception, configuration change directions, etc. of the management information of an SNMP protocol which were described previously may be performed through IF of host computer to the channel IF section 402. At this invention, the monitor mechanism section 404 are collected at a fixed interval by the COP section 403 in the access logical volume number and its operating frequency for every channel collected by the channel IF section 402. This collected data is brought together in the SVP section 412 through the monitor mechanism section 404. Therefore, the information on the access frequency for the monitor information total means 405. The monitor mechanism section 404 may be realized by whichever of the control every channel path number in all the unit disk controllers 401 will gather for the SVP section 412. Here, the monitor mechanism section 404 may be realized by whichever of the control every channel path number in all the unit disk controllers 401 will gather for the SVP section 412. Here, the monitor mechanism section 404 may be realized as a channel IF section. Program performed on a processor, and the hardware mounted as a channel IF section. [0018] Thus, information is notified or displayed on a system operator from a judgment and the result of the information which the monitoring information from each channel was totaled and totaled in the SVP section 412 by the flow 414 of the communication with a system operator 411 and a notice and a signal input dashed line in drawing. It shall replace with a system operator 411 as the information notified or displayed on a system operator is shown in drawing with a thick dashed line, the SVP section shall be performed between high order equipment 406. In this case, as the information notified or displayed on a system operator is shown in drawing with a thick dashed line, the SVP section 412 has the interface 416 with high order equipment, and exchanges information between high order equipment 406. Of course, in addition to this, to say nothing of the ability to take the various roots, there is no need that especially the physical mounting position of IF which exchanges management information has a place limited in this invention.

[0019] Drawing 5 is a system configuration view for explaining more concretely the total of the access data between the host computer by the SVP (maintenance terminal) section 412, and a disk and shows access between a host computer and a logical volume typically. While host computers 500, —, 50N are combined with a disk controller 560 through the SAN switch 520, the unit disk controllers 510, —, 51N are constituted by the disk controller 560. While logical volume #000+N is constituted in each unit disk controller, it is coordinated with the common cross coupling network 530 between unit disk controllers. Here, it was displayed in the unit disk controller of a disk controller that a logical volume was constituted because a disk unit could necessarily be seen from a host computer and the logical volume used as the unit of access were seen.

... and connect the SAN switch 520

connected with the SAN switch [0022]. It seems that the mapping information on the path of the mapping table managed by SAN manager 540 is shown in drawing 6 at this time. Although it is the display which paid its attention only to the logical path explained by drawing 5 in drawing 6, the correspondence relation between the path number of all the logical paths with which this mapping table combines a host side path number (physical path) and disk side path number (physical path), and the logical path on a chart has become.

local volume number accessed by this logical volume. [0023] Drawing 7 is drawing showing the example of a display which showed the result of the access total frequency to the logical volume which totaled at the maintenance terminal 550 containing the SVP (maintenance terminal) section 412 which explained access which was explained by drawing 5 and drawing 6 by drawing 4, and the accessed frequency for every channel path number to the logical volume numerical order. For example, it is totaled as shown in a table 706, and it is displayed on the display 707 as the output section of the maintenance terminal 550, and an operator is notified. The number of all the logical volumes from which a table 706 constitutes the period of a total unit in 701, and constitutes a system in 702 the total table 706 constitutes the period of a total unit in 701, and constitutes a system in 702 the access frequency [as opposed to each logical volume to 703] (a lead (R) →) The access frequency which goes a light (W) exception via the channel path in the same unit disk controller of the aforementioned total access frequency in 704 Access from the channel of other unit disk controllers, i.e., the access frequency which goes via the common cross coupling network 530, is shown in 705, respectively. In addition, the channel path said by 705 will be a path between the SAN switch 520 and a 511 to 51 n unit disk controller, if it says in the example of drawing 5. [0024] Here, when it sees about the access frequency of a lead (R) of logical volume #000000 to access from the channel of other unit disk controllers which go via channel path #N-1 and #N to the access frequency which goes via the channel path in the same unit disk controller having only 200X103, i.e., the access frequency which goes via the common cross coupling network 530, is as large as 1200X103. Therefore, when this is considered by the system configuration of drawing 5 and 000000 is logical volume LVO#000 in the logical volume # unit disk controller 511 in drawing 5, access from the channel of other unit disk controllers which go via MUBU this logical volume / the empty logical volume in 51n of unit disk controllers], copying then

[0025] The example of the message 708 for notifying a system operator 411 of the message channel path #N-1, and #N can be improved.

shall be performed by pushing the click of the **left** button 710 by carrying out an inquiry pair, and button 709 for directions, and the "No (N)" button 710 by carrying out an inquiry pair, and according to pointing devices, such as a mouse, or "Y" of a keyboard.

[0022] In a second, more detailed embodiment, the copy of a logical volume is performed although the system was not constructed until it saw so that all host computers may understand a logical volume table by referring to, as for whether it is an opening, it is natural that this table is also updated corresponding to this. This table is good also as that with which a host computer is equipped, or a disk controller may be equipped with it. Of course, you may prepare for the SVP section 412.

[0023] When it will have not only MUBU or the copy of a logical volume but the SAN switch 52 again in order to improve access to a logical volume if drawing 5 is referred to, it turns out that an access path can be changed by change of this and it can improve. That is, access to logical volume LVOL#001 of unit disk controllers through the common cross coupling network FIG. 5 can avoid going via the common cross coupling

network 530 by the change of the SAN switch 520 like access to logical volume 511 of unit 510 of unit disk controllers through the SAN switch 520 with the path 53 of a host computer 500.

[0028] What it will lead to the improvement of access that replace with the common cross coupling network 550 the logical path 72 which accesses logical volume LY01#001 of 511 of unit disk controllers through the common cross coupling network 530 with the path 52 of a host computer 500 like a logical path 13, and it considers as SAN switch 520 course is easily understood so that it may turn out easily that each logical path shown by the thick line is understood.

specifically compared with [drawing 5](#). Hereafter, the case where an access path is changed is explained to the SAN switch 520.

[0029] Hereafter, the case where an access path is changed is explained to the SAN switch 520 between a disk controller 500 and a 500 n host computer.

[0030] With reference to [drawing 8](#), the change of the SAN switch 520 is the same if 500 and a 500 n host computer is explained concretely. In addition, [drawing 8](#) is the same if the matter about the change of [drawing 5](#) and the SAN switch 520 is removed.

[0031] In [drawing 5](#), as stated also in advance, access to logical volume LV01#001 of 51n of unit disk controllers through the common cross coupling network 530 with the path 52 of a host computer 500 is performed by the logical path 72 which goes via the common cross coupling network 530. As is illustrated to [drawing 7](#), it is totaled at the maintenance terminal (SVP) 550, and this is displayed on a system operator as INFORMATION as shown in [drawing 9](#) by being judged together with the frequency of access. In this case, by [drawing 5](#), although access to logical volume LV01#001 of unit disk controller #n is changed into the logical path of SAN switch 520 course so that [drawing 8](#) may show, since all the paths from the SAN switch 520 to switch 520 course so that [drawing 8](#) may show, since all the paths from the SAN switch 520 to unit disk controller #n are what is used, it recommends extending path #N+1. Of course, if there is a non-used path, naturally, using this will be recommended. When a system operator performs recommendation to this recommendation, path #N+1 is extended between the SAN switch 520 and unit disk controller #n so that it may mention later. Then, the "Yes (Y)" button 903 is clicked with pointing devices, such as a mouse, or "Y" of a keyboard is pushed. By this, the SAN manager 540 can recognize that path #N+1 was extended between the SAN switch 520 and unit disk controller #n, can newly set up a logical path 79, and can change access to logical volume LV01#001 of unit disk controller #n from a path 52 into the logical path of SAN switch 520 course. When not accepting the "No (N)" button 902 is clicked with pointing devices, such as a "N" of a keyboard is pushed.

g., and drawing 406, in the meaning which understood that processing was completed to this. In addition, the method of making the messages that it made such a change after 1203 in the meaning which understood that performing a altogether change of the path explained until now by a report to the high order equipment (a switch, host computer) 406 in drawing 4, and the function of the high order equipment 406 according to this, without sending the message to recommend ex post facto

approval, and displaying or notifying it may be used. [0035] Next, the case where MUBU or the copy of a logical volume is carried out is explained. Drawing 13 is drawing showing the example of access between the host computer in the system which consists of situations except the SAN switch 520, and the logical volume in each unit disk controller from the system explained by drawing 5. The 500 to 50 n host computer is connected to unit disk controller #1 - #N (511 to 51 n) of a disk controller 500 using path #0 - path #N (511 to 56). A path 51 by the logical path 131 as the path between the path which a host computer uses, and the logical volume of an access place is shown all over drawing 13. Logical volume LVOL#001 of the unit disk controller 511, A path 52 minds the common cross coupling network 530 by the logical path 132, to logical volume LVOL#000 of 51 n of unit disk controllers A path 53 minds the common cross coupling network 530 by the logical path 133. to logical volume LVOL#001 of 51 n of unit disk controllers A path 54 minds the common cross coupling network 530 by the logical volume LVOL#002 of the unit disk controller 511. The path 55 shows the state where have accessed logical volume LVOL#000 of 51 n of unit disk controllers by the logical path 135, and the path 56 has accessed logical volume LVOL#000 of the unit disk controllers by the logical path 136, respectively. At this time, the total result of an access situation as shown by drawing 14 is obtained, this information is set in the SVP section 550, and a system operator is displayed notified of it, or a host is notified of it. In addition, the logical volume number in drawing 14 lets the logical volume number of all 511 to 51 n unit disk controllers pass. It follows, for example, logical volume LVOL#000 of the unit disk controller 511 correspond to the logical volume 00000000 of drawing 14, and logical volume LVOL#N of 51 n of unit disk controllers corresponds to logical volume N of drawing 14. The meaning which the contents of the information shown by drawing 14 have is important here. [0036] As resolution of this information, making between 00:00~23:00 per day into a total unit is the frequency using a cross coupling network 530. and the example of arbitrary periods and shown by distinction of a lead (R) and light access (W), and the example of logical path accesses is only drawing. From this information, if the logical volume which a certain logical path accesses it is good for the unit disk controller with which the frequency which uses a cross coupling network most becomes small MUBU [a subordinate / a logical volume]. And when the lead which leads the common cross coupling network 530, the logical volume to which is also carried out through the common cross coupling network will be carried out small and which a certain logical path accesses it is good for the unit disk controller with which the frequency which uses a cross coupling network, most becomes small MUBU [a subordinate / a logical volume]. And when situations, like the lead demand from a certain path surely increases rapidly in a time zone processing by copying volume etc. is possible only for the time zone.

[0037] It is good to copy the logical volume LVOL000 which is 51 n of unit disk controllers which a logical path 132 accesses to logical volume LYOL00K of the unit disk controller 511 so that it can say from the access frequency of a logical path having been shown by the thick line in the example of drawing 13. Moreover, it is good MUBU [the logical volume LVOL002 of the unit disk controller 511 which a logical path 134 accesses / logical volume LVOL00M of 51 n of unit disk controllers]. Here, naturally the logical volume of a copy or the MUBU point is in an empty state. Moreover, MUBU [this example / the logical volume LVOL000 of 51 n of unit disk controllers] shows an example of the message 1501 notified to high order equipment 406, because the logical path 135 has accessed this.

[0038] Drawing 15 shows an example of the total result of drawing 4 as it was displayed on SVP550 or drawing 4 explained according to the total result of drawing 14. In this example, since the number of a logical volume and the number of the logical volume of the unit disk controller of drawing 13 which are expressed by drawing 14 did not correspond clearly, it expresses like logical volumes C and D, —, E. If it says by correspondence with drawing 13, logical volume C in drawing 15 corresponds to the logical volume LVOL00M of 51 n of unit disk controllers, logical volume D corresponds to logical volume LVOL000 of 51 n of unit disk controllers, and logical volume F corresponds to logical volume LYOL00K of the unit disk controller 511.

http://www4.ipdlipo.go.jp/cgi-bin/tran_web.cgi.cgi
2003/06/10

volume LVOL002 of the unit disk controller 511, and copying the logical volume LVOL000 of 51 n of unit disk controllers to logical volume LVOL00M of 51 n of unit disk controllers] To INFORMATION as a recommended. [logical volume LVOL00M of 51 n of unit disk controllers]

system operator indicated to be to drawing 15 about operation of MUBU or a copy at this time in performing recommendation, the "Yes (Y)" buttons 1503 and 1504 are clicked with pointing devices, such as a mouse, or it pushes "Y" of a keyboard. In not performing recommendation, the "No (N)" buttons 1502 and 1505 are clicked with pointing devices, such as a mouse, or it pushes "N" of a keyboard. It is good that what is necessary is just made to operate it corresponding to an execution wizard being started and the order of the real way being automatically shown by this corresponding to this operation.

[0040] Probably operation will be performed by directions of the direct control to the high order equipment 406 shown in drawing 4, as previously explained to change or extension of a logical path, and it will be clear that it is good also as what reports only a result to an operator, when result to the operator by such MUBU or a copy.

[0041] Drawing 16 is the example of INFORMATION1601 which reports a result, after the processing according to directions is made, as a result of operating the "Yes (Y)" buttons 1503 and 1504 in drawing 15. this — moreover, it is also the example of a display of the operation using the common cross coupling network in this invention of a logical volume. To the case where the copy or MUBU of a logical volume which access concentrates is performed between unit disk controllers, it can carry out efficiently as follows. For example, suppose that one of the logical volumes which access which minds a common cross coupling network from many host computers concentrates is in one of the unit disk controllers by MUBU or the copy using the common cross coupling network in this invention of a logical volume. To the case of the logical volume which access concentrates copies the logical volume concerned to the logical volume of the opening of the unit disk controller to a useful unit disk controller, the burden of a common cross coupling network is reduced, and since all unit disk controllers are combined through the common cross coupling network in this invention, even if a data transfer band is small, it can do with what can fully respond.

[0043] For example, suppose that one of the logical volumes which access concentrates is in the coupling network 1713 to the unit disk controllers 1701 and 1702 to the unit disk controllers 1701-1702. Then, a copy 1710 is performed from the unit disk controllers 1701 and 1702 to the unit disk controllers 1703 and 1704. Consequently, the number with the content of the logical volume which access concentrates of logical volumes increases to 4 from 2. Subsequently, further, if a copy 1712 is performed the number with the content of the logical volume which access concentrates volumes will increase from these logical volumes to 8 from 4. That is, when copying volume to N sets of unit disk controllers, the copy between N-1 time of a unit disk controller is performed. The notice access to a logical volume is received according to the monitor mechanism and the notice mechanism which it explained with reference to drawing 4. If a load distribution is performed by copying the logical volume on which is straddled and access concentrates a unit disk controller for every unit disk controller As for the common cross coupling network which connects between unit disk controllers between a unit disk controller and a unit disk controller, the band to the number twice of a unit disk controller of the data transfer rate between the channel-interface section inside a unit disk controller and the disk interface section, and the cache memory section, it is enough.

[0044] Drawing 18 is drawing having shown the outline at the time of considering the common cross coupling network which combines between unit disk controllers as a simple interconnection path. By this invention, since there is a path which combines between unit disk controllers between unit disk controllers in at least two or more physically different paths, as shown in drawing, the number of access paths of the common cross coupling network 1805 increases the unit disk controllers 1801-1804 most in the case where it connects mutually with two paths.

http://www4.ipdlipo.go.jp/cgi-bin/tran_web.cgi.cgi

2003/06/10

respectively, in the example of the simplest cross coupling network. [0045] Drawing 19 is an example by which the number of paths decreases the case and the contrary of drawing 18 most, and between unit disk controllers is combined with them in two or more physically different paths. The unit disk controllers 1901-1904 are combined with the connection path 1907, respectively, for example, two of the paths 2 shown in the paths 1 and 1906 from the unit disk controller 1901 to the unit disk controller 1903 which show the path of access by 1905 in drawing exist. In such connection, the number of paths decreases most (however, a bus connection removes.).

[0046] Drawing 20 is the ** type view showing an example of the power supply redundant configuration of the whole disk controller in this invention. Here, the common cross coupling network 2005 performs electric path control. When that is not right, the power supply itself is unnecessary. The case where every two power supplies of 2006-2013 are supplied is shown to the unit disk controllers 2001-2004 and the common cross coupling network 2005, respectively. Thus, when using a redundant power supply for each simply, the number of power supplies increases most.

[0047] Drawing 21 shows the example which redundancy-izes a power supply using the power supplies 2006-2013 used for the unit disk controllers 2001-2004 to connect, respectively about the common cross coupling network 2005, in order to reduce the number of power supplies to drawing 20.

[0048] Drawing 22 is drawing showing an example of the mounting method of the redundant power supply in composition like drawing 21. Drawing 22 is an example which four unit disk controllers 2202-2205 are mounted in one console 2201, two are arranged up and down among these (drawing 22 (A)), and the structure of these 2 tiering is made to face each other in respect of the back plane of each unit disk controller, and is arranged (drawing 22 (B)). It is shown in drawing 22 (A) -- as --- the unit disk controller 2202-2205 -- respectively -- channel IF ****, disk IF ****, cache memory ****, and shared memory **** -- having (referring to drawing 1 - drawing 3) --- it has two power supplies These elements are mounted on the back plane 2212-2215. Although the console 2201 was equipped with common cross coupling **** 2208 and illustration was omitted, as drawing 1 - drawing 3 explained using the connector 2209 of common cross coupling **** 2208, between each unit disk controller is combined. Moreover, since some terminals 2209 of this connector are used as a power supply introduction terminal of common cross coupling **** 2208, a power supply is supplied in parallel through a cable 2215 from the power supply of the unit disk controller 2202-2205. Especially the physical relationship of each part grade here does not have a meaning. Moreover, it is a book, even if it connects each with a cable using a back plane all over drawing, although common cross coupling **** was connected with each unit disk controller.

[0049] [Effect of the Invention] By this invention, when two or more unit disk controllers function as

one disk controller with a common cross coupling network, cost is held down and the equipment with which the effect of the number of a unit disk controller is reflected effective in a performance can be offered.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] Drawing in which this invention shows the outline of an example of the conventional composition of the target computer system.

[Drawing 2] Drawing showing the outline of the example of composition when this invention attains the increase in efficiency of a system configuration according to the SAN environment to the conventional disk array of the target computer system.

[Drawing 3] Drawing showing the outline of the conventional disk array of composition of that this invention connected the small disk controller of the conventional scale of the target computer system with the common cross coupling network.

[Drawing 4] Drawing showing the outline of the logical volume access frequency monitor mechanism for enabling service of this invention.

[Drawing 5] Drawing showing the system configuration for explaining more concretely the total of the access data between the host computer by the SVP (maintenance terminal) section, and a disk.

[Drawing 6] Drawing showing an example of the mapping information on the path of the mapping table managed by the SAN manager in drawing 5.

[Drawing 7] Drawing showing the example of a display which showed the result of the access total frequency to a logical volume which totaled at the maintenance terminal containing the SVP (maintenance terminal) section, and the accessed frequency for every channel path number to the logical volume numerical order.

[Drawing 8] Drawing explaining the example which changes a logical path with the SAN switch between a disk controller and a host computer.

[Drawing 9] Drawing showing an example of INFORMATION to the system operator offered by this invention corresponding to access to the logical volume of the unit disk controller through the common cross coupling network of a host computer.

[Drawing 10] Drawing showing other examples of INFORMATION to the system operator offered by this invention.

[Drawing 11] Drawing showing the example of a display of the response to the result which carried out "Yes (Y)" button grabbing which accepts this to INFORMATION shown in drawing 7, drawing 9, and drawing 10.

[Drawing 12] Drawing showing the example of a display of the response to the result which carried out "No (N)" button grabbing which does not accept this to INFORMATION shown in drawing 7, drawing 9, and drawing 10.

[Drawing 13] Drawing showing the example of access between the host computer which brings a result which carries out MUBU or the copy of a logical volume, and the logical volume in each unit disk controller.

[Drawing 14] Drawing showing an example of the access situation total result in drawing 13.

[Drawing 15] Drawing showing the example of INFORMATION which recommends the path change according to the total result of drawing 14.

[Drawing 16] Drawing showing the example of INFORMATION which reports a result after the processing according to directions is made, as a result of operating the "Yes (Y)" button which

accepts INFORMATION in drawing 15 .

[Drawing 17] Drawing explaining the outline of the load distribution by MUBU or the copy using the cross coupling network in this invention of a logical volume.

[Drawing 18] Drawing showing the outline at the time of considering the common cross coupling network which combines between unit disk controllers as a simple interconnection path.

[Drawing 19] Drawing showing the example by which the number of paths decreases the case and the contrary of drawing 18 most, and between unit disk controllers is combined with them in two or more physically different paths.

[Drawing 20] The ** type view showing an example of the power supply redundant configuration of the whole disk controller in this invention.

[Drawing 21] It is drawing showing the example which redundancy-izes a power supply about a common cross coupling network using the ***** power supply for unit disk controllers which connects in order to reduce the number of power supplies to drawing 20 .

[Drawing 22] Drawing showing an example of the mounting method of the redundant power supply of composition of being shown in drawing 21 .

[Description of Notations]

101, 500, a 50n:host computer, 102 : A magnetic disk unit, 103: The channel-interface section, 104 : The disk interface section, 105: A cross coupling network, a 106:cross coupling network, 107 : The cache memory section, 108: The shared memory section, 109,315,560 : A disk controller, 212 : A magnetic-tape-storage and 309:unit disk controller 210,301:common cross coupling network, 412: The SVP (maintenance terminal) section, 401:unit disk controller, 402 : The channel IF section, 403: The COP section, the 404:monitor mechanism section, 405 : A monitor information total means, 406: The configuration control section, the 407:output section, the 408:input section, 409 : Notice/display means, 410: A directions input means, a 411:system operator, 414 : The flow of communication of information, 415 : A means to transmit information, a 520:SAN switch, 510, —, 51n:unit disk controller, 530: A common cross-coupling network, a LVOL:logical volume, a 520:SAN switch, 51 and 52, —, 56, 61 and 62, —, 66:physics path, 71 and 72, —, 75, 131, 132, —, 136: A logical path, a 540:SAN manager.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is an example of the block diagram showing the outline of the disk controller concerning this invention.

[Drawing 2] It is an example of the block diagram showing the cache control information of the disk controller concerning this invention.

[Drawing 3] It is the flow chart showing an example of the whole operation concerning this invention.

[Drawing 4] It is the flow chart showing an example of the updating access demand processing concerning this invention.

[Drawing 5] It is the flow chart showing an example of the coherence processing concerning this invention.

[Drawing 6] It is the flow chart showing an example of the coherence processing concerning this invention.

[Drawing 7] It is the flow chart showing an example of the reference access demand processing concerning this invention.

[Drawing 8] It is the flow chart showing an example of the reference access demand processing concerning this invention.

[Drawing 9] It is the flow chart showing an example of the cache management method concerning this invention.

[Drawing 10] It is the flow chart showing an example of the reference access demand processing concerning this invention.

[Drawing 11] It is an example of the block diagram showing the cache of the disk controller concerning this invention.

[Drawing 12] It is an example of the block diagram showing the cache of the disk controller concerning this invention.

[Drawing 13] It is the flow chart showing an example of the cache management method concerning this invention.

[Drawing 14] They are other examples of the block diagram showing the outline of the disk controller concerning this invention.

[Drawing 15] They are other examples of the block diagram showing the outline of the disk controller concerning this invention.

[Drawing 16] It is the flow chart showing an example of the data configuration method of the disk controller concerning this invention.

[Drawing 17] It is the block diagram showing the outline of the conventional disk controller concerning this invention.

[Drawing 18] It is the block diagram showing the outline of the conventional disk controller concerning this invention.

[Description of Notations]
 0 — ... — a host computer and 1 — ... — a disk subsystem and 2 — ... — a channel and 3 — ... — control information and 4 — ... — storing data and 5 — ... — reference data and 6 — ... — updating data and 7 — ... — a logic disk and 10 — ... — a disk controller and 11 — ... — the

channel-control section and 12 — ... — a control memory and 13 — ... — the cache memory
section and 14 — ... — a disk control section and 15 — ... — a joint

[Translation done.]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2002-333956
(P2002-333956A)

(43)公開日 平成14年11月22日(2002.11.22)

(51) Int.Cl.⁷

識別記号
304
540

F I
G 06 F 3/06

テ-マコ-ト(参考)

304N 5B065
540

審査請求 未請求 請求項の数12 OL (全 14 頁)

(21) 出願番号 特願2001-138424(P2001-138424)

(22)出願日 平成13年5月9日(2001.5.9)

(71) 出願人 000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目 6 番地

(72) 発明者 藤林 昭
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(72) 発明者 藤本 和久
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(74) 代理人 100068504
弁理士 小川 勝男 (外 2 名)

最終頁に続く

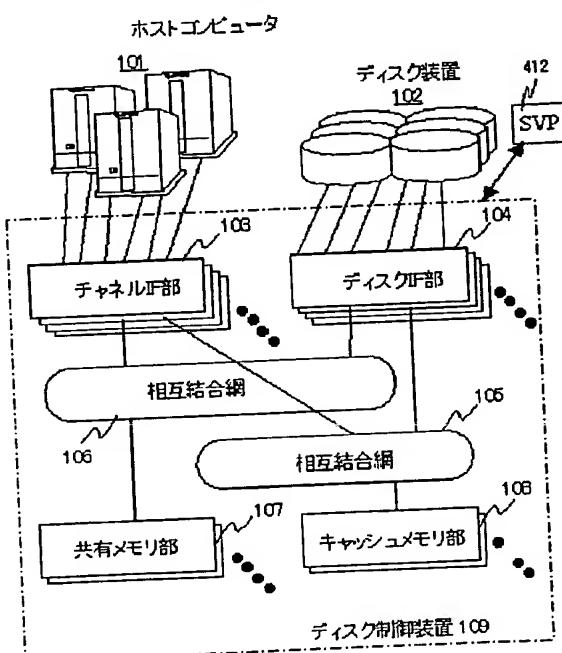
(54) 【発明の名称】 ディスク制御装置を用いた計算機システムおよびその運用サービス

(57) 【要約】

【課題】 単位ディスク制御装置間を渡るアクセスがある場合にも十分に性能を出す為には相互結合網の帯域を非常に大きくしなければならないが、その為にコストが高くなる。

【解決手段】 論理ボリュームのアクセス頻度をモニタし、上位装置にアクセスバスの変更を促したり、論理ボリュームのムーブまたはコピーを各単位ディスク制御装置に持つことで、相互結合網は主として論理ボリュームのコピーの為に利用されるようにすることで、必要な帯域を低く抑える。

图 1



【特許請求の範囲】

【請求項1】 ホストコンピュータと、該ホストコンピュータとのアクセスを制御する為の一つないし複数のプロセッサを有するチャネルインターフェース部と、磁気ディスク装置と、該磁気ディスク装置とのアクセスを制御する為の一つないし複数のプロセッサを有するディスクインターフェース部と、前記ホストコンピュータおよび磁気ディスク装置間で書込／読出するデータを格納するキャッシュメモリ部および前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間を相互に接続する機能を有する相互結合網と、よりなる構成を単位とする単位コンピュータシステムが前記相互結合網を介して接続する共通相互結合網により接続されたコンピュータシステムであって、前記単位コンピュータシステムのホストコンピュータからの前記共通相互結合網を介しての前記磁気ディスク装置への論理バスによるアクセス状況を監視する手段と、該アクセス状況を集計し所定の周期あるいはタイミングでアクセス頻度状況を保守端末に表示させ、あるいは、システム管理者に通知する手段をコンピュータシステム内に有することを特徴とするコンピュータシステム。

【請求項2】 ホストコンピュータと、該ホストコンピュータとのアクセスを制御する為の一つないし複数のプロセッサを有するチャネルインターフェース部と、磁気ディスク装置と、該磁気ディスク装置とのアクセスを制御する為の一つないし複数のプロセッサを有するディスクインターフェース部と、前記ホストコンピュータおよび磁気ディスク装置間で書込／読出するデータを格納するキャッシュメモリ部および前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間を相互に接続する機能を有する相互結合網と、よりなる構成を単位とする単位コンピュータシステムが前記ホストコンピュータとチャネルインターフェース部とを接続する手段の間で共通相互結合網により接続されたコンピュータシステムであって、前記単位コンピュータシステムのホストコンピュータからの前記共通相互結合網を介しての前記磁気ディスク装置への論理バスによるアクセス状況を監視する手段と、該アクセス状況を集計し所定の周期あるいはタイミングでアクセス頻度状況を保守端末に表示させ、あるいは、システム管理者に通知する手段をコンピュータシステム内に有することを特徴とするコンピュータシステム。

【請求項3】 ホストコンピュータと、該ホストコンピュータとのアクセスを制御する為の一つないし複数のプロセッサを有するチャネルインターフェース部と、磁気ディスク装置と、該磁気ディスク装置とのアクセスを制御する為の一つないし複数のプロセッサを有するディスクインターフェース部と、前記ホストコンピュータおよび磁気ディスク装置間で書込／読出するデータを格納する

キャッシュメモリ部および前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間を相互に接続する機能を有する相互結合網と、よりなる構成を単位とする単位コンピュータシステムが前記相互結合網を介して接続する第1の共通相互結合網により接続されるとともに、前記ホストコンピュータとチャネルインターフェース部とを接続する手段の間で第2の共通相互結合網により接続されたコンピュータシステムであって、前記単位コンピュータシステムのホストコンピュータからの前記第1および第2の共通相互結合網を介しての前記磁気ディスク装置への論理バスによるアクセス状況を監視する手段と、該アクセス状況を集計し所定の周期あるいはタイミングでアクセス頻度状況を保守端末に表示させ、あるいは、システム管理者に通知する手段をコンピュータシステム内に有することを特徴とするコンピュータシステム。

【請求項4】 前記共通相互結合網を介しての磁気ディスク装置の論理ボリュームと論理バスとのアクセス頻度状況を保守端末に表示させ、あるいは、システム管理者に通知する手段に代えて、上位装置に報告する手段を有する請求項1ないし3のいずれかに記載のコンピュータシステム。

【請求項5】 ホストコンピュータと、該ホストコンピュータとのアクセスを制御する為の一つないし複数のプロセッサを有するチャネルインターフェース部と、磁気ディスク装置と、該磁気ディスク装置とのアクセスを制御する為の一つないし複数のプロセッサを有するディスクインターフェース部と、前記ホストコンピュータおよび磁気ディスク装置間で書込／読出するデータを格納するキャッシュメモリ部および前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間を相互に接続する機能を有する相互結合網と、よりなる構成を単位とする単位コンピュータシステムが前記相互結合網を介して接続する共通相互結合網により接続されたコンピュータシステムに適用され、前記単位コンピュータシステムのホストコンピュータからの前記共通相互結合網を介しての前記磁気ディスク装置への論理バスによるアクセス状況を監視するとともに該アクセス状況を集計し所定の周期あるいはタイミングでアクセス頻度状況に応じてシステム管理者または上位装置に対し、論理バスとディスク側バスのマッピング変更を提案することを特徴とするコンピュータシステムの運用サービス。

【請求項6】 ホストコンピュータと、該ホストコンピュータとのアクセスを制御する為の一つないし複数のプロセッサを有するチャネルインターフェース部と、磁気ディスク装置と、該磁気ディスク装置とのアクセスを制御する為の一つないし複数のプロセッサを有するディスクインターフェース部と、前記ホストコンピュータおよび磁気ディスク装置間で書込／読出するデータを格納する

キャッシュメモリ部および前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間を相互に接続する機能を有する相互結合網と、よりなる構成を単位とする単位コンピュータシステムが前記ホストコンピュータとチャネルインターフェース部とを接続する手段の間で共通相互結合網により接続されたコンピュータシステムに適用され、前記単位コンピュータシステムのホストコンピュータからの前記共通相互結合網を介しての前記磁気ディスク装置への論理バスによるアクセス状況を監視するとともに該アクセス状況を集計し所定の周期あるいはタイミングでアクセス頻度状況に応じてシステム管理者または上位装置に対し、論理バスとディスク側バスのマッピング変更を提案することを特徴とするコンピュータシステムの運用サービス。

【請求項7】ホストコンピュータと、該ホストコンピュータとのアクセスを制御する為の一つないし複数のプロセッサを有するチャネルインターフェース部と、磁気ディスク装置と、該磁気ディスク装置とのアクセスを制御する為の一つないし複数のプロセッサを有するディスクインターフェース部と、前記ホストコンピュータおよび磁気ディスク装置間で書込／読出するデータを格納するキャッシュメモリ部および前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間を相互に接続する機能を有する相互結合網と、よりなる構成を単位とする単位コンピュータシステムが前記相互結合網を介して接続する第1の共通相互結合網により接続されるとともに、前記ホストコンピュータとチャネルインターフェース部とを接続する手段の間で第2の共通相互結合網により接続されたコンピュータシステムに適用され、前記単位コンピュータシステムのホストコンピュータからの前記共通相互結合網を介しての前記磁気ディスク装置への論理バスによるアクセス状況を監視するとともに該アクセス状況を集計し所定の周期あるいはタイミングでアクセス頻度状況に応じてシステム管理者または上位装置に対し、論理バスとディスク側バスのマッピング変更を提案することを特徴とするコンピュータシステムの運用サービス。

【請求項8】前記論理バスとディスク側バスのマッピング変更に代えてホストコンピュータからの磁気ディスク装置のアクセス先論理ボリュームのコピーおよび／またはムーブを提案する請求項5ないし7のいずれかに記載のコンピュータシステムの運用サービス。

【請求項9】前記提案に対応して、上位装置において、前記ディスク制御装置よりの論理バスとディスク側バスのマッピング変更および／またはアクセス先論理ボリュームのコピーおよび／またはムーブの指示の内容に従って、当該論理バスマッピングの変更および／またはアクセス先論理ボリュームのコピーおよび／またはムーブを自動的に実行する論理バスとアクセス先論理ボリューム

の対応を管理する手段を備える請求項5ないし8のいずれかに記載のコンピュータシステム。

【請求項10】前記提案とともに、前記ディスク制御装置よりの論理バスとディスク側バスのマッピング変更および／またはアクセス先論理ボリュームのコピーおよび／またはムーブの指示の内容に従って、当該論理バスマッピングの変更および／またはアクセス先論理ボリュームのコピーおよび／またはムーブを実行するための手順自動指示手段を起動させるかどうかを同時に表示および／または通知する請求項5ないし8のいずれかに記載のコンピュータシステム。

【請求項11】前記単位コンピュータシステム間のそれぞれが独立した複数の接続線よりなる共通相互結合網により接続された請求項1ないし3のいずれかに記載のコンピュータシステム。

【請求項12】前記単位コンピュータシステムのそれぞれが独立した二つの電源により駆動されるとともに、前記共通相互結合網は前記独立した二つの電源を統合した電源によりにより駆動される請求項1ないし3のいずれかに記載のコンピュータシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、データを複数のディスク装置に格納するディスクアレイのディスク制御装置を用いる計算機システムおよびその運用サービスに関する。

【0002】

【従来の技術】現在、多量のデータを扱うところでは、データセンタなどでメインフレーム等と接続する大型ディスクアレイ、テープバックアップ装置、オープン系のサーバー群に接続する小中型ディスクアレイなどさまざまなストレージ機器が導入され、それぞれに情報を格納している。現在はこれらの格納情報が各自に有機的に強く連携しているとは言い難い。そこで、SAN (Storage Area Network) という概念が導入され、さまざまなストレージ機器をネットワークで接続し、情報を管理しようと言う動きが活発になって来た。大規模なデータセンタでも今まで大型ディスクアレイで高信頼、高性能、高機能なストレージシステムを構築していたが、SANにより、小中型のディスクアレイも加わって、今まで以上に大規模なストレージシステムが構築されつつある。

【0003】一方、今後は、小中型ディスクアレイにおいても高性能、高信頼が要求されてくる。そこで、現在のディスクアレイでは、SOHOなどの小規模から銀行などの大規模なストレージシステムをスケーラブルに支えるディスクアレイが必要となってくるとともに、これを効果的に運用可能にするサービスが必要になってくる。

【0004】従来のディスクアレイは図1に示すようなディスク制御装置を中心とするものである。ホストコン

ピュータ101とディスク制御装置109間のデータ転送を実行する複数のチャネルインターフェース（以下チャネルIF）部103と磁気ディスク装置102とディスク制御装置109間のデータ転送を実行する複数のディスクインターフェース（以下ディスクIF）部104と、チャネルIFとディスクチャネルIF間で読み・書きされるデータを格納するキャッシュメモリ部107とディスクアレイ制御装置109に関する制御情報（例えば、チャネルIF部103とキャッシュメモリ部107とのデータ転送に関する制御情報等）を格納する共有メモリ部108を備え、各チャネルIF部103、ディスクIF部104とキャッシュメモリ部107は相互結合網105で接続され、各チャネルIF部103、ディスクIF部104と共有メモリ部108も相互結合網106で接続される。ここで言う相互結合網は、スイッチ、ループ、バスなど、あらゆる接続手段を意味する。ここで、412はSVP（保守端末）部であり、後述するようにホストコンピュータ101とディスク制御装置109との間の情報伝送のチャネルバス番号毎のアクセス頻度の情報を集めるものである。

【0005】このような構成により一つのディスク制御装置を形成する場合、スケーラビリティとは、最小構成時を基本装置として、それに対するコンポーネントの追加、つまりオプションの拡張用コンポーネントを逐次追加していく形式である。つまり最大構成までのオプションコンポーネントを追加する為の拡張機構を最小構成時から有している必要があり、小規模構成の時でも拡張時に必要となる機構が備わっており、それらは、基本構成で運用している場合には不要の機構である。装置コストは基本構成時点では必然的に割高になると言える。また、ホストインターフェースの高速化、コネクティビティの向上（接続可能ホストインターフェース数の増加）に対応していく為には、当該の拡張コンポーネント向け接続機構も高速化、拡張性の向上（拡張可能コンポーネント数の増加）に対応する為に、さらに高コストになり、基本構成時に割高になる可能性が大きい。

【0006】これに対し、図2に概要を示すように、ディスクアレイを用いたSAN（Storage Area Network）環境によりシステム構成の効率化を図ることが行われている。ホストコンピュータ群101は共通相互結合網210を通じて、ディスク制御装置109に接続している。ディスク制御装置は複数のチャネルIF部103と複数のディスクIF部104と複数の共有メモリ部107と複数のキャッシュメモリ部108とが、それぞれ相互結合網105、106で結合した構成になっており、ディスクIF部は複数のディスク装置102と接続している。ディスク制御装置109とそれに接続するディスク装置102はディスクアレイとして機能する。また共通相互結合網210は多種のストレージデバイスを接続可能であり、磁気テープ記憶装置212等も接続され

る。具体的にはファイバチャネルスイッチやループ、LAN等あらゆるネットワークが考えられる。このような形態の場合、例えば、小型ディスクアレイ装置を数十台、数百台と接続して、多数の論理ボリュームの集合体として、上位のホストコンピュータにシステムを見せることが可能であり、従来型の高可用、高信頼の大型ディスクアレイと同等の大容量と図中の経路211のようにディスク制御装置間の論理ボリュームのコピーなどの高機能が、実現できる。しかし、小型ディスクアレイ装置は、大型のディスクアレイのような高可用性、高信頼性を追及したものではないという問題がある。しかし、安価な小型ディスクアレイの集合体であるという価格面でのメリットもある。ここでも、412はSVP（保守端末）部であり、ホストコンピュータ101とディスク制御装置109との間の情報伝送のチャネルバス番号毎のアクセス頻度の情報を集めるものである。

【0007】さらに、類似の構成として、図3に示す様に、ホストコンピュータ群101とディスク装置群102とをそれぞれ接続するチャネルIF部103、ディスクIF部104とを共有メモリ部107とキャッシュメモリ部108と接続する相互結合網105、106からなり、ディスクアレイとして機能する図1、図2で示したディスク制御装置よりも規模の小さいディスク制御装置を単位ディスク制御装置309として、該単位ディスク制御装置を複数台、共通相互結合網310で接続し、全体としてディスクアレイを構成することもできる。この場合、単位ディスク制御装置は従来型ディスク制御装置の1/4～1/2程度の規模とすることで、実装もコンパクトになることで、コスト低下が見込める。また、中心となる共通相互結合網310にも必要最小限の帯域を用意することで、装置全体のコストを低下できる。ここでも、412はSVP（保守端末）部であり、ホストコンピュータ101とディスク制御装置315、より厳密には全ての単位ディスク制御装置309との間の情報伝送のチャネルバス番号毎のアクセス頻度の情報を集めるものである。

【0008】さらに、図示は省略したが、図2と図3に示すシステム構成を統合したシステムも在りうる。すなわち、図3におけるホストコンピュータ101間を図2に示す共通相互結合網210で結合する。そうすると、ホストコンピュータが自己に直結している単位ディスク制御装置を介さずに、他のホストコンピュータに直結している単位ディスク制御装置にアクセスすることができる。この結果、単位ディスク制御装置間の相互結合網を経由する必要が無いから、アクセスが改善できる。

【0009】

【発明が解決しようとする課題】このように、小規模構成でディスクアレイ装置として機能する単位ディスク制御装置を相互結合網により接続して、大規模構成までの

スケーラビリティを実現することができ、拡張時の機構を予め装置内に構成しておくことが不要となり、初期コストが低減される。しかし、ホストコンピュータとディスクとの関係で見ると制御装置間相互結合網についてのバスが偏る可能性があり、このことがシステムの効率的な運用を阻害する可能性がある。勿論、単位ディスク制御装置間相互結合網に必要な帯域を大きくとることが一つの解決策にはなるが、特定のバスの偏りに備えて大きな帯域を準備することは、単位ディスク制御装置を相互結合網により接続するシステムとしたことによる初期コストの低減のメリットを低減させてしまうことになる。

【0010】さらには、制御装置間相互結合網についてのバスと電源の冗長化を考慮する必要があるが、これも、単純に制御装置間相互結合網用の冗長バスおよび冗長電源を用意すると、単純にコスト増加となる。

【0011】本発明の目的は上記の構造を持ったディスク制御装置での単位ディスク制御装置間相互結合網に必要な帯域を低減できるように、単位ディスク制御装置間のデータ転送を可能な限り低減可能としたコンピュータシステムおよびその運用サービスを提供することにあり、さらに必要なら、該相互結合網への電力供給源の適な冗長化を提案するものである。

【0012】

【課題を解決するための手段】上記目的を達成するため、本発明では、ホストコンピュータの利用するアクセスとそのアクセス先ボリュームが同一の単位ディスク制御装置内に存在する確率を向上する為、アクセス状況を監視し、採取したアクセス情報により、接続する上位装置に対して、最適なバスを使用するよう推奨したり、単位ディスク制御装置間渡りを頻発している当該論理ボリュームのムーブもしくはコピーを行うことを推奨したりする情報を、システム管理者に保守端末や管理者用web端末などを通じるなどして、表示または通知する。また、それらを自動的に実行させることも可能とする。このようにして、単位ディスク制御装置間のデータのやり取りはボリュームのコピーやムーブを主とするよう装置全体を制御することで、相互結合網に必要な帯域を抑える。

【0013】また、共通相互結合網への電力供給の冗長化も、共通相互結合網は2台以上の単位ディスク制御装置が存在する場合にのみ必要なことから、複数の単位ディスク制御装置の電源を用いて、冗長電源とすることで、不要な電源の増加をおさえる。

【0014】

【発明の実施の形態】本発明の提供するディスク制御装置とその実装方法の実施例について、以下に図面を示し実施例を参照して詳細に説明する。

【0015】図4は本発明のサービスを可能にするための論理ボリュームアクセス頻度モニタ機構の概要を示す図である。ただし、ここでは、システムオペレータ、ス

イッチなどの上位装置、ホストコンピュータへの通知手段や表示手段および指示入力手段を特定せずに説明する。実施時には、CRTを用いて、表示したり、ネットワークを通じてオペレータの端末にメールで通知を行つたり、オペレータの端末にブラウザを通じて表示する。

【0016】さらには、ディスク制御装置とスイッチやホストコンピュータなどを接続するインターフェースとして考えられるファイバチャネルやギガイーサネット

(登録商標)などには、SNMPのような構成管理プロトコルも用意されており、これを利用すればディスク制御装置と上位装置との間で管理情報の相互伝達が可能となる。指示入力手段としてはキーボードを使った入力手段、ネットワークを利用してブラウザベースでの入力手段など、様々な方法が適用できる。

【0017】SVP(保守端末)部412は、構成制御部406、出力部407および入力部408を備え、ディスクアレイの装置全体の構成情報管理や制御を行う。SVP(保守端末)部412とシステムオペレータ411との間には、通知/表示手段409および指示入力手段410が備えられ、SVP(保守端末)部412の出力部407の信号をシステムオペレータ411に通知/表示を行うことができるとともにシステムオペレータ411からの指示を入力部408に受け付けることができる。構成制御部406はディスクアレイからの構成情報を受けたり、ディスクアレイに構成変更の指示を行うものである。また、チャネルIF部402に対して、先に述べたSNMPプロトコルの管理情報の送受信や構成変更指示等を、ホストコンピュータのIFを通じて実行するよう、互いに情報を伝達する手段415を備える。本発明では、モニタ機構部404により収集されるチャネル毎のアクセス論理ボリューム番号とその使用頻度を、チャネルIF部402中の制御プロセッサ部403により一定間隔で収集する。この収集されたデータは、モニタ情報集計手段405を通じてSVP部412に集められる。したがって、SVP部412には全単位ディスク制御装置401におけるチャネルバス番号毎のアクセス頻度の情報が集まることになる。ここで、モニタ機構部404はプロセッサ上で実行される制御プログラムと、チャネルIF部として実装するハードウェアとのどちらで実現されていても良い。

【0018】このようにして、各チャネルからのモニタリング情報が集計され、図中の破線で示す情報伝達の流れ414によって、SVP部412において集計した情報の判定とそこからシステムオペレータに情報が通知または表示される。システムオペレータ411に代えて、上位装置406との間で通知および信号入力を行うものとすることもできる。この場合は、システムオペレータに通知または表示される情報は、図に太い破線で示すように、SVP部412が、上位装置とのインターフェース416を持って上位装置406との間で情報交換を行

うものとなる。もちろん、この他にも、種々のルートを取りうることは言うまでもなく、本発明において、管理情報をやり取りする IF の物理的な実装位置は特に場所を限定される必要は無い。

【0019】図5は、SVP（保守端末）部412によるホストコンピュータとディスク間のアクセスデータの集計をより具体的に説明するためのシステム構成図であり、ホストコンピュータと論理ボリュームとの間のアクセスを模式的に示したものである。ホストコンピュータ500、---、50nがSANスイッチ520を介してディスク制御装置560に結合されるとともに、ディスク制御装置560には単位ディスク制御装置510、---、51nが構成されている。各単位ディスク制御装置内には論理ボリューム#000---#Nが構成されるとともに、単位ディスク制御装置間は共通相互結合網530により連繋される。ここで、ディスク制御装置の単位ディスク制御装置内に論理ボリュームが構成されるように表示したのは、ホストコンピュータからはディスク装置が見えるわけではなく、アクセスの単位となる論理ボリュームが見えるに過ぎないからである。

【0020】ホストコンピュータと論理ボリュームとの間のアクセスを各論理バスについて見ると以下のようである。単位ディスク制御装置#0の論理ボリュームLVOL#000はバス51、SANスイッチ520、バス61を経由する論理バス71を利用してアクセスされる。単位ディスク制御装置#nの論理ボリュームLVOL#001はバス52、SANスイッチ520、バス63、共通相互結合網530を経由する論理バス72を利用してアクセスされる。単位ディスク制御装置#nの論理ボリュームLVOL#000はバス54、SANスイッチ520、バス65を経由する論理バス74を利用してアクセスされる。単位ディスク制御装置#nの論理ボリュームLVOL#002はバス56、SANスイッチ520、バス66を経由する論理バス75を利用してアクセスされる。さらに、単位ディスク制御装置#nの論理ボリュームLVOL#003はバス53、SANスイッチ520、バス64を経由する論理バス73を利用してアクセスされる。

【0021】複数のホストコンピュータ500---50nとディスク制御装置560が、SANスイッチ520を介して接続されているが、SANスイッチ520には、ホストコンピュータがSANスイッチ配下の装置の物理的な接続状況を意識することが必要ないようにする為に、ホスト側バス#0---#N 51---56とディスク側バス#0---#N 61---66との組み合わせを、論理アクセスバスとして管理し、そのマッピングテーブルはSANマネージャ540で管理するものとされる。SANマネージャの実態は管理ソフトウェアであり、SANスイッチ520上のどこか（ホストコンピュータ含む）に実装される。本発明においては、SANマネージャ54

0の実装される位置を問題にしない。専用のハードウェアがSANスイッチ520と接続されるものであっても良い。

【0022】この時、SANマネージャ540で管理されるマッピングテーブルのバスのマッピング情報は図6に示すようなものとなっている。図6では、図5で説明した論理バスにのみ着目した表示となっているが、このマッピングテーブルはホスト側バス番号（物理バス）とディスク側バス番号（物理バス）とを結合する全ての論理バスのバス番号と、この論理バスによりアクセスされる論理ボリューム番号との対応関係が一覧表となっているものである。

【0023】図7は、図5、図6で説明したようなアクセスを、図4で説明したSVP（保守端末）部412を含む保守端末550で集計した論理ボリュームに対するアクセス総頻度とアクセスを行ったチャネルバス番号ごとの頻度の結果を論理ボリューム番号順に示した表示例を示す図である。たとえば、テーブル706のように集計されて保守端末550の出力部としてのディスプレ707に表示されオペレータに通知される。テーブル706は、701に集計単位の期間を、702にシステムを構成する全ての論理ボリュームの番号を、703にそれぞれの論理ボリュームに対する総アクセス頻度（リード（R）、ライト（W）別）を、704に前記総アクセス頻度の内の同一単位ディスク制御装置内チャネルバスを経由してのアクセス頻度を、705に他の単位ディスク制御装置のチャネルからのアクセス、すなわち、共通相互結合網530を経由してのアクセス頻度を、それぞれ示す。なお、705で言うチャネルバスは、図5の例でいえば、SANスイッチ520と単位ディスク制御装置511---51nとの間のバスである。

【0024】ここで、たとえば、論理ボリューム#000000のリード（R）のアクセス頻度について見ると、同一単位ディスク制御装置内チャネルバスを経由してのアクセス頻度が 200×10^3 しかないのに対して、チャネルバス#N-1および#Nを経由する他の単位ディスク制御装置のチャネルからのアクセス、すなわち、共通相互結合網530を経由してのアクセス頻度は 1200×10^3 と大きい。したがって、このことを図5のシステム構成で考えてみると、論理ボリューム#000000が図5における単位ディスク制御装置511内の論理ボリュームLVOL#000である場合には、この論理ボリュームを単位ディスク制御装置51n内の空き論理ボリュームにムーブするかコピーすることとすれば、チャネルバス#N-1および#Nを経由する他の単位ディスク制御装置のチャネルからのアクセスを改善することができる。

【0025】図7の下段部には、このようなアクセスの改善に関するメッセージをシステムオペレータ411に通知するためのメッセージ708の例を示す。ここで

は、一般的に論理ボリュームAのコピーを、論理ボリュームBにコピーすることを推奨する旨のメッセージを保守端末550の出力部としてのディスプレー707上に表示したものとした。表示のフォーマットはウインドウを用いたGUI(グラフィックユーザインターフェイス)であるものとした。この通知はアクセスモニタリングの集計結果を表示すると共に、論理ボリュームコピー元論理ボリューム番号Nを単位ディスク制御装置番号N以下の論理ボリュームにコピーすることを推奨する旨とそのボリュームコピーの実行手順自動指示機構(いわゆるウィザード)の起動をするかどうかの問い合わせと併せ持つメッセージである。したがって、その問い合わせに対して、実行を指示用の「Yes(Y)」ボタン709および「No(N)」ボタン710の表示を行い、マウス等のポインティングデバイスによる「Yes(Y)」のクリックもしくはキーボードの「Y」を押すことで、ボリュームコピーの手順自動指示機構が実行されるものとすることができます。

【0026】なお、論理ボリュームがどのようにホストコンピュータに割り当てられ、あるいは、空きであるか否かは、全てのホストコンピュータが論理ボリュームテーブルを参照することで分かるようシステムが構成されていることは言うまでも無いが、論理ボリュームのムーブあるいはコピーが行われたときなどには、このテーブルもこれに対応して更新されるのは当然である。このテーブルはホストコンピュータに備えられるものとしても良く、あるいは、ディスク制御装置に備えられるものであっても良い。もちろん、SVP部412に備えられても良い。

【0027】再び、図5を参照すると、論理ボリュームへのアクセスを改善するには、論理ボリュームのムーブあるいはコピーのみならず、SANスイッチ520を備えるときは、これの切替えによりアクセスバスの変更をして改善することができる事が分かる。すなわち、ホストコンピュータ500のバス52による共通相互結合網530を介しての単位ディスク制御装置51nの論理ボリュームLVOL#001へのアクセスは、SANスイッチ520の切替えにより、ホストコンピュータ500のバス53によるSANスイッチ520を介しての単位ディスク制御装置51nの論理ボリュームLVOL#003へのアクセスと同様に、共通相互結合網530を経由することを回避できる。

【0028】具体的には、図5に太線で示す各論理バスを比較して見ると容易に分かるように、ホストコンピュータ500のバス52による共通相互結合網530を介して単位ディスク制御装置51nの論理ボリュームLVOL#001にアクセスする論理バス72は、論理バス73のように、共通相互結合網530に代えてSANスイッチ520経由とすることがアクセスの改善につながるであろうことは容易に理解される。

【0029】以下、ディスク制御装置560とホストコンピュータ500-50nの間にあるSANスイッチ520に対して、アクセスバスの変更を行う場合について説明する。

【0030】図8を参照して、ディスク制御装置560とホストコンピュータ500-50nの間にあるSANスイッチ520の切替えを具体的に説明する。なお、図8は、図5とSANスイッチ520の切替えに関する事項を除けば同じである。

【0031】図5では、先にも述べたように、ホストコンピュータ500のバス52による共通相互結合網530を介しての単位ディスク制御装置51nの論理ボリュームLVOL#001へのアクセスが共通相互結合網530を経由する論理バス72により行われている。このことが、図7に例示したように保守端末(SVP)550で集計され、アクセスの頻度と合わせ判定されることにより、図9に示すようなINFORMATIONとしてシステムオペレータに表示される。この場合には、図8から分かるように、単位ディスク制御装置#nの論理ボリュームLVOL#001へのアクセスをSANスイッチ520経由の論理バスに変更するものであるが、図5ではSANスイッチ520から単位ディスク制御装置#nへのバスは全て使用されているものとなっているから、バス#N+1を増設することを推奨するものとなる。もちろん、不使用のバスがあれば、これを使用することを推奨することになることは当然である。この推奨に対してシステムオペレータが推奨を実行することとした場合には、後述するように、SANスイッチ520と単位ディスク制御装置#nとの間にバス#N+1を増設する。その後、「Yes(Y)」ボタン903を、マウス等のポインティングデバイスによりクリックし、あるいは、キーボードの「Y」を押す。このことにより、SANマネージャ540はSANスイッチ520と単位ディスク制御装置#nとの間にバス#N+1が増設されたことを認知することができ、新たに論理バス79を設定してバス52から単位ディスク制御装置#nの論理ボリュームLVOL#001へのアクセスをSANスイッチ520経由の論理バスに変更することができる。受け入れないときには、「No(N)」ボタン902を、マウス等のポインティングデバイスによりクリックし、あるいは、キーボードの「N」を押す。

【0032】なお、この例では、バス#N+1が増設されることが必要であったので、このような手順としたが、空きのバスが在り、これへの変更が示唆されたときは、「Yes(Y)」ボタン903を、マウス等のポインティングデバイスによりクリックし、あるいは、キーボードの「Y」を押すだけで良い。図10は、空きバスへの変更を示唆するINFORMATIONの例である。バスが増設されることが必要である場合には、先に増設をしておかないと、SANマネージャ540で管理

されるマッピングテーブルのバスのマッピング情報のみが変更されるとホストコンピュータのアクセスが誤ったものとなる可能性があるからである。単なる変更であるときは、SANマネージャ540がバスの変更を認識すれば良いだけであるから、図10のような空きバスへの変更を示唆するINFORMATIONに対しては、これを受け入れるなら、「Yes (Y)」ボタン1003を、マウス等のポインティングデバイスによりクリックし、あるいは、キーボードの「Y」を押すだけで良い。受け入れないときには、「No (N)」ボタン1002を、マウス等のポインティングデバイスによりクリックし、あるいは、キーボードの「N」を押す。

【0033】図11および図12は、それぞれ、図7、図9および図10に示されるINFORMATIONに対して、これを受け入れる「Yes (Y)」ボタン709、903および1003を、および受け入れない「No (N)」ボタン710、902および1002をマウス等のポインティングデバイスによりクリックしたときの保守端末550の出力部としてのディスプレー707上の応答の表示例である。これに対しては処理が完了したことを了解した意味で「Yes (Y)」ボタン1103および1203を押せばよい。

【0034】なお、今まで説明したバスの変更は、推奨するメッセージなどを出さずに、図4における上位装置（スイッチ、ホストコンピュータ）406への報告とこれに応じた上位装置406の機能によりすべて実行してから、そのような変更を行ったというメッセージを事後承諾として表示または通知する方法でも構わない。

【0035】次に、論理ボリュームのムーブまたはコピーをする場合について説明する。図13は、図5で説明したシステムからSANスイッチ520を除いた状況で構成されているシステムにおけるホストコンピュータと各単位ディスク制御装置内の論理ボリュームとの間のアクセスの例を示す図である。ホストコンピュータ500-50nがバス#0～バス#N(51-56)を使ってディスク制御装置560の単位ディスク制御装置#1～#N(511-51n)に接続されている。ホストコンピュータが使用するバスとアクセス先の論理ボリュームとの関係は図中に示される通り、バス51は論理バス131により単位ディスク制御装置511の論理ボリュームLVOL#000と、バス52は論理バス132により共通相互結合網530を介して単位ディスク制御装置51nの論理ボリュームLVOL#000に、バス53は論理バス133により共通相互結合網530を介して単位ディスク制御装置51nの論理ボリュームLVOL#000に、バス54は論理バス134により共通相互結合網530を介して単位ディスク制御装置511の論理ボリュームLVOL#002に、バス55は論理バス135により単位ディスク制御装置51nの論理ボリュームLVOL#000に、バス56は論理バス136に

より単位ディスク制御装置51nの論理ボリュームLVOL#002に、それぞれ、アクセスしている状態を示している。この時、図14で示されるようなアクセス状況の集計結果が得られ、該情報はSVP部550において、システムオペレータに表示または通知されるか、ホストに通知される。なお、図14における論理ボリューム番号は全ての単位ディスク制御装置511-51nの論理ボリューム番号を通してのものである。したがって、たとえば、単位ディスク制御装置511の論理ボリュームLVOL#000が図14の論理ボリューム000000に対応し、単位ディスク制御装置51nの論理ボリュームLVOL#Nが図14の論理ボリュームNに対応する。ここで重要なのは、図14で示される情報の中身の持つ意味である。

【0036】本情報の分解能として、リード(R)、ライトアクセス(W)の区別と任意の期間、図の例では一日当たりの19.00-23.00の間、を集計単位とすることが示されている。本情報より、ある論理バスのアクセスする論理ボリュームが共通相互結合網530を通過してのリードのみであれば、ホストコンピュータに直結する単位ディスク制御装置に当該論理ボリュームのコピーを持たせることが、相互結合網を利用する頻度を小さくすることになり、また、ある論理バスのアクセスする論理ボリュームが共通相互結合網530を通じてリードライトもされている場合は、最も相互結合網を利用する頻度が小さくなる単位ディスク制御装置に配下に論理ボリュームをムーブするのが良い。そして、集計期間に応じて、例えば深夜、早朝時間帯に必ず、あるバスからのリード要求が急激に多くなるなどの状況を把握した場合は、その時間帯のみボリュームのコピーを行うなどの細やかな処理が可能である。

【0037】図13の例では、論理バスのアクセス頻度を太線で示したことから言えるように、論理バス132がアクセスする単位ディスク制御装置51nの論理ボリュームLVOL#000を、単位ディスク制御装置511の論理ボリュームLVOL#00Kにコピーするのが良い。また、論理バス134がアクセスする単位ディスク制御装置511の論理ボリュームLVOL#002を単位ディスク制御装置51nの論理ボリュームLVOL#00Mにムーブするのが良い。ここで、コピーあるいはムーブ先の論理ボリュームが空きの状態であることは当然である。また、この例で、単位ディスク制御装置51nの論理ボリュームLVOL#000をムーブしないのは、これに論理バス135がアクセスしているからである。

【0038】図15は、図14の集計結果に応じて、SVP550に表示されるか、図4で説明したように、上位装置406に通知されるメッセージ1501の一例を示している。この例では、図14で表現されている論理ボリュームの番号と図13の単位ディスク制御装置の論理ボリュームの番号とが明確に対応しないので、論理ボ

リュウムC, D, ---, Eのように表現した。図13との対応で言えば、図15における論理ボリュームCは単位ディスク制御装置511の論理ボリュームLVOL002に対応し、論理ボリュームDは単位ディスク制御装置51nの論理ボリュームLVOL00Mに対応し、論理ボリュームEは単位ディスク制御装置51nの論理ボリュームLVOL000に対応し、論理ボリュームFは単位ディスク制御装置511の論理ボリュームLVOL00Kに対応するものである。

【0039】この対応を示す図14の集計結果より、先ず、単位ディスク制御装置511の論理ボリュームLVOL002を単位ディスク制御装置51nの論理ボリュームLVOL00Mにムーブし、単位ディスク制御装置51nの論理ボリュームLVOL000を単位ディスク制御装置511の論理ボリュームLVOL00Kにコピーすることが推奨されるのである。このとき、ムーブまたはコピーの操作については、システムオペレータは図15に示されるようなINFORMATIONに対しても、推奨を実行することとした場合には、「Yes (Y)」ボタン1503, 1504をマウス等のポイントティングデバイスによりクリックし、あるいは、キーボードの「Y」を押す。推奨を実行しないこととした場合には、「No (N)」ボタン1502, 1505をマウス等のポイントティングデバイスによりクリックし、あるいは、キーボードの「N」を押す。この操作に対応して、実行ウィザードが起動され、これにより実行手順が自動的に示されるのに対応して操作をすれば良いようになるのが良い。

【0040】このようなムーブまたはコピーを伴う場合も、先に、論理バスの変更あるいは増設に対して説明したように、図4に示す上位装置406に対する直接操作の指示により操作が行われて、オペレータには結果のみを報告するものとしても良いことは明らかであろう。

【0041】図16は、図15における「Yes (Y)」ボタン1503, 1504が操作された結果指示に応じた処理がなされた後に結果を報告するINFO RMATION1601の例である。これは、また、上記の直接操作の指示によるオペレータへの操作結果の表示例でもある。システムオペレータは、了解の意味で、「Yes (Y)」ボタン1603, 1604を操作すればよい。

【0042】図17は、本発明における共通相互結合網を利用した論理ボリュームのムーブもしくはコピーによる負荷分散の概要を説明する図である。アクセスの集中する論理ボリュームのコピーもしくはムーブが単位ディスク制御装置間で行われる場合に対しても以下のように効率的に行うことができる。例えば、多くのホストコンピュータから共通相互結合網を介してのアクセスが集中する論理ボリュームの一つが単位ディスク制御装置の一つにあるとする。このような場合、本発明では、すべて

の単位ディスク制御装置が共通相互結合網を介して結合されているから、アクセスが集中する論理ボリュームのコピーを持つことが有用な単位ディスク制御装置に対してその単位ディスク制御装置の空きの論理ボリュームに対して当該論理ボリュームのコピーをすれば、共通相互結合網の負担を低減してデータ転送帯域が小さくても十分に対応できるものとできる。

【0043】たとえば、アクセスが集中する論理ボリュームの一つが単位ディスク制御装置1701にあるとする。この場合、まず、単位ディスク制御装置1701から1702に、共通相互結合網1713を介して、コピー1709を実行する。その後、単位ディスク制御装置1701, 1702から単位ディスク制御装置1703, 1704にコピー1710を実行する。その結果、アクセスが集中する論理ボリュームの内容を持った論理ボリュームの数は2から4に増加する。ついで、これらの論理ボリュームから、さらに、コピー1712を実行すれば、アクセスが集中する論理ボリュームの内容を持った論理ボリュームの数は4から8に増加する。つまり、N台の単位ディスク制御装置にボリュームをコピーする場合、N-1回の単位ディスク制御装置間のコピーが行われる。つまり、図4を参照して説明したモニタ機構と通知機構により論理ボリュームへのアクセスに対して、単位ディスク制御装置を跨ってアクセスの集中する論理ボリュームを単位ディスク制御装置ごとにコピーすることで負荷分散を行えば、単位ディスク制御装置と単位ディスク制御装置の間で単位ディスク制御装置間を接続する共通相互結合網は専ら論理ボリュームのコピー作成の為にその帯域が使用されることになり、そのデータ転送帯域は単位ディスク制御装置内部のチャネルインターフェース部およびディスクインターフェース部とキャッシュメモリ部間のデータ転送速度の単位ディスク制御装置台数倍と同等であれば十分である。

【0044】図18は単位ディスク制御装置間を結合する共通相互結合網を単純な相互接続バスとした場合の概略を示した図である。本発明では、単位ディスク制御装置間は少なくとも2つ以上の物理的に異なる経路で、単位ディスク制御装置間を結合するバスがあるので、もつとも単純な相互結合網の例は、図のように単位ディスク制御装置1801～1804をそれぞれ、2つのバスで互いに接続した場合で、共通相互結合網1805は最もアクセスバス数が多くなる。

【0045】図19は、図18の場合と反対に、もっともバス数が少くなり、かつ、2つ以上の物理的に異なる経路で単位ディスク制御装置間が結合されている例である。単位ディスク制御装置1901～1904は接続バス1907でそれぞれ結合されており、例えば、単位ディスク制御装置1901から単位ディスク制御装置1903へのアクセスの経路は図中の1905で示す経路1と1906で示す経路2の2つが存在する。このよう

な接続の場合もっともバス数が少なくなる（但し、バス接続は除く。）。

【0046】図20は本発明におけるディスク制御装置全体の電源冗長構成の一例を示す模式図である。ここで、共通相互結合網2005は電気的な経路制御を行うもの（スイッチ、ハブ、その他）である。そうでない場合は電源そのものが不要である。単位ディスク制御装置2001～2004および共通相互結合網2005に対してそれぞれ、2006～2013の電源が2つづつ供給される場合を示している。このように、単純にそれぞれに冗長電源を用いるとした時は、もっとも電源数が多くなる。

【0047】図21は、図20に対して電源数を低減するため、共通相互結合網2005については、接続する単位ディスク制御装置2001～2004にそれぞれ用いられる電源2006～2013を用いて電源を冗長化する例を示す。

【0048】図22は、図21のような構成の場合の冗長電源の実装方法の一例を示す図である。図22は一つのコンソール2201に四つの単位ディスク制御装置2202～2205が実装されており、これらの内二つが上下に配列され（図22（A））、この二段積みの構造が各単位ディスク制御装置のバックプレーンの面で向き合わせて配置されている（図22（B））例である。図22（A）に示すように、単位ディスク制御装置2202～2205は、それぞれ、チャネルIF部群、ディスクIF部群、キャッシュメモリ部群および共有メモリ部群を備える（図1～図3参照）とともに、二つの電源を備える。これらの要素はバックプレーン2212～2215上に実装されている。コンソール2201には共通相互結合網部2208が備えられ、図示は省略したが、共通相互結合網部2208のコネクタ2209を利用して図1～図3で説明したように各単位ディスク制御装置間が結合される。また、このコネクタの一部の端子2209は共通相互結合網部2208の電源導入端子として利用されるので、単位ディスク制御装置2202～2205の電源からケーブル2215を介して、並列に電源が供給される。ここでの各部位の位置関係は特に意味を持つものではない。また図中ではケーブルにより、各単位ディスク制御装置と共通相互結合網部を接続したが、バックプレーンを用いて各々を接続しても本発明の効果は変わらない。

【0049】

【発明の効果】本発明により、共通相互結合網により、複数の単位ディスク制御装置が一つのディスク制御装置として機能する場合に、コストを抑え、単位ディスク制御装置の台数の効果が性能に有効に反映される装置が提供できる。

【図面の簡単な説明】

【図1】本発明が対象とするコンピュータシステムの從

来の構成の一例の概要を示す図。

【図2】本発明が対象とするコンピュータシステムの従来のディスクアレイにSAN環境によりシステム構成の効率化を図った場合の構成例の概要を示す図。

【図3】本発明が対象とするコンピュータシステムの従来の規模の小さいディスク制御装置を共通相互結合網で接続した構成の従来のディスクアレイの概要を示す図。

【図4】本発明のサービスを可能にするための論理ボリュームアクセス頻度モニタ機構の概要を示す図。

【図5】SVP（保守端末）部によるホストコンピュータとディスク間のアクセスデータの集計をより具体的に説明するためのシステム構成を示す図。

【図6】図5におけるSANマネージャで管理されるマッピングテーブルのバスのマッピング情報の一例を示す図。

【図7】SVP（保守端末）部を含む保守端末で集計した、論理ボリュームに対するアクセス総頻度とアクセスを行ったチャネルバス番号ごとの頻度の結果を論理ボリューム番号順に示した表示例を示す図。

【図8】ディスク制御装置とホストコンピュータの間にあるSANスイッチにより論理バスを切替える例を説明する図。

【図9】ホストコンピュータの共通相互結合網を介しての単位ディスク制御装置の論理ボリュームへのアクセスに対応して、本発明により提供されるシステムオペレータへのINFORMATIONの一例を示す図。

【図10】本発明により提供されるシステムオペレータへのINFORMATIONの他の一例を示す図。

【図11】図7、図9および図10に示されるINFO R MATIONに対して、これを受け入れる「Yes (Y)」ボタン操作した結果に対する応答の表示例を示す図。

【図12】図7、図9および図10に示されるINFO R MATIONに対して、これを受け入れない「No (N)」ボタン操作した結果に対する応答の表示例を示す図。

【図13】論理ボリュームのムーブまたはコピーをする結果となるホストコンピュータと各単位ディスク制御装置内の論理ボリュームとの間のアクセスの例を示す図。

【図14】図13におけるアクセス状況集計結果の一例を示す図。

【図15】図14の集計結果に応じたバス変更を推奨するINFORMATIONの例を示す図。

【図16】図15におけるINFORMATIONを受け入れる「Yes (Y)」ボタンが操作された結果指示に応じた処理がなされた後に結果を報告するINFORMATIONの例を示す図。

【図17】本発明における相互結合網を利用した論理ボリュームのムーブもしくはコピーによる負荷分散の概要を説明する図。

【図18】単位ディスク制御装置間を結合する共通相互結合網を単純な相互接続バスとした場合の概略を示す図。

【図19】図18の場合と反対に、もっともバス数が少くなり、かつ、2つ以上の物理的に異なる経路で単位ディスク制御装置間が結合されている例を示す図。

【図20】本発明におけるディスク制御装置全体の電源冗長構成の一例を示す模式図。

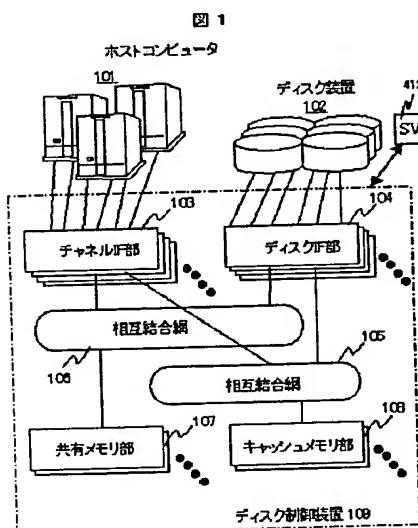
【図21】図20に対して電源数を低減するために、共通相互結合網については、接続する単位ディスク制御装置用いられる電源を用いて電源を冗長化する例を示す図。

【図22】図21に示す構成の冗長電源の実装方法の一例を示す図。

【符号の説明】

101, 500, 50n: ホストコンピュータ、102: 磁気ディスク装置、103: チャネルインターフェース部、104: ディスクインターフェース部、105: 相互結合網、106: 共通相互結合網、107: キャッシュメモリ部、108: 共有メモリ部、109: ディスク制御装置、110: ディスク装置、111: ハードディスク、112: SVP (保守端末) 部、113: チャネルIF部、114: 制御プロセッサ部、115: モニタ情報集計手段、116: 構成制御部、117: 出力部、118: 入力部、119: 通知/表示手段、120: 指示入力手段、121: システムオペレータ、122: 情報伝達の流れ、123: 情報を伝達する手段、124: SANスイッチ、125: SANマネージャ、126: 共通相互結合網、127: 論理ボリューム、128: SANスイッチ、129: SANマネージャ、130: 物理バス、131: SANスイッチ、132: SANマネージャ、133: 物理バス、134: SANスイッチ、135: SANマネージャ、136: 論理バス、137: SANマネージャ。

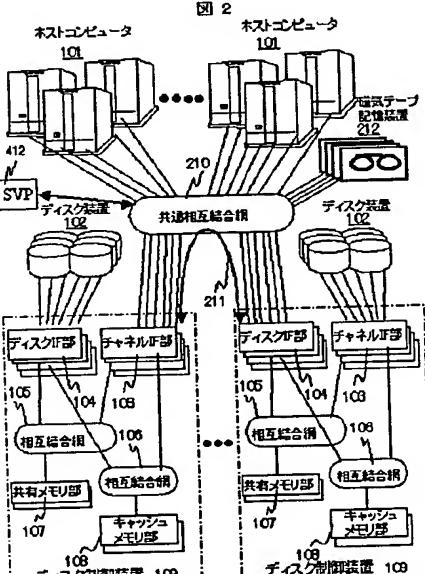
【図1】



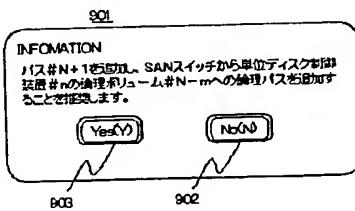
【図6】

アクセスバス番号	ホストバス	ディスクバス	論理ボリューム番号
71	61	61	00000000単位ディスク装置#N-mのLVOL000が対応
72	62	63	0N-m単位ディスク装置#N-mのLVOL001が対応
73	63	64	0N-m+2単位ディスク装置#N-mのLVOL003が対応
74	64	65	0N-m-1単位ディスク装置#N-mのLVOL000が対応
75	65	66	0N-m+1単位ディスク装置#N-mのLVOL002が対応

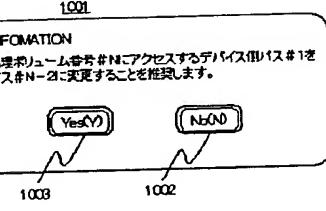
【図2】



【図9】

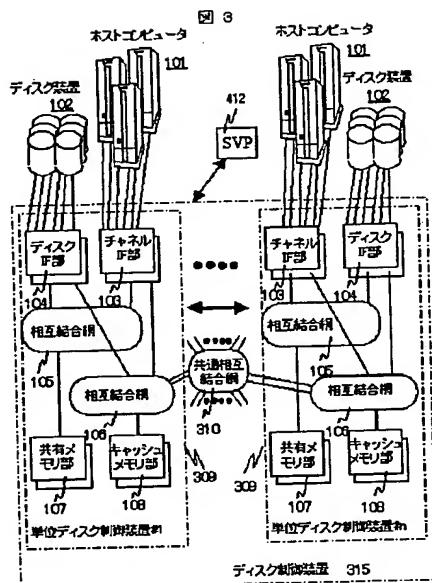


【図10】

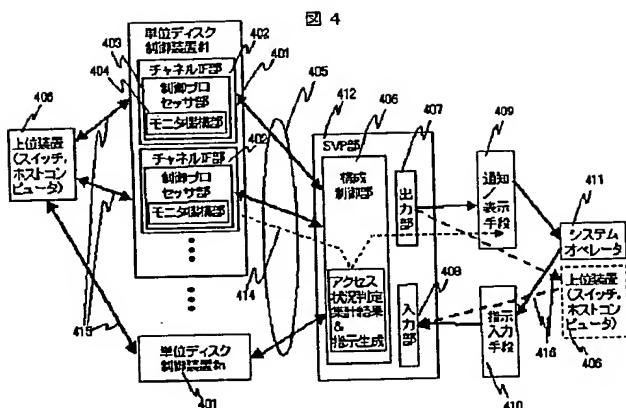


5: 相互結合網、106: 相互結合網、107: キャッシュメモリ部、108: 共有メモリ部、109, 31: 5, 60: ディスク制御装置、212: 磁気テープ記憶装置、309: 単位ディスク制御装置210, 301: 共通相互結合網、412: SVP (保守端末) 部、401: 単位ディスク制御装置、402: チャネルIF部、403: 制御プロセッサ部、404: モニタ機構部、405: モニタ情報集計手段、406: 構成制御部、407: 出力部、408: 入力部、409: 通知/表示手段、410: 指示入力手段、411: システムオペレータ、414: 情報伝達の流れ、415: 情報を伝達する手段、520: SANスイッチ、510, ---, 51n: 単位ディスク制御装置、530: 共通相互結合網、LVOL: 論理ボリューム、520: SANスイッチ、51, 52, ---, 56, 61, 62, ---, 66: 物理バス、71, 72, ---, 75, 131, 132, ---, 136: 論理バス、540: SANマネージャ。

【图3】

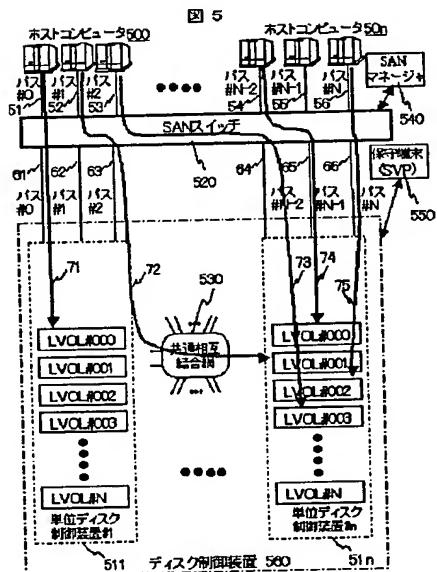


【図4】



【图8】

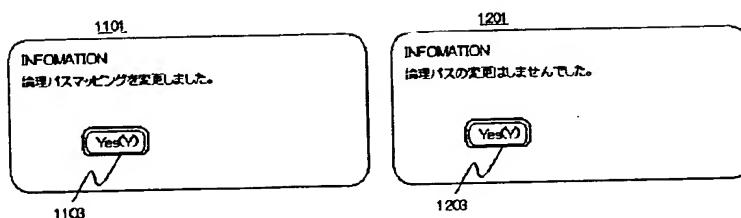
【図5】



[図 1 1]

[図 12]

11



【図 7】

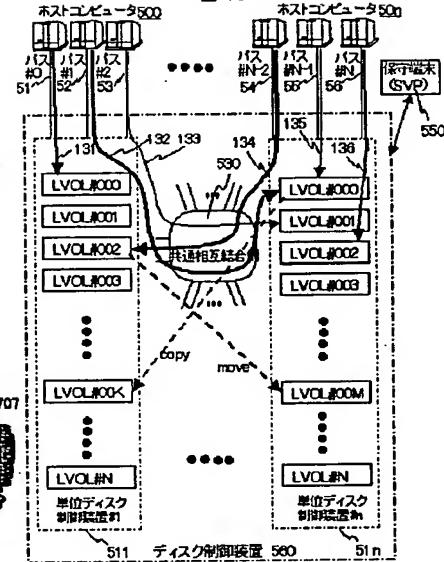
図 7

統計単位 : 1WEEK		論理ボリューム 番号(LVOL#)	読み書き頻度 ($\times 10^3$)	同一単位ディスク制御 装置内チャネルバス頻度 ($\times 10^3$)	他単位ディスク制御装置チャネルバス 頻度($\times 10^3$)	#00	#01	#N-1	#N
# 0000000	R 8000	200		120 200		1200	1200			
	W 2000	800		200 200		200	200			
# 0000001	R 5000	2000		400				400	400	
	W 5000	2000		1000 1000						
# 0000002	R 8000	2400		1000				1000		
	W 2000	600		100			300			
# 0000003	R 5000	2500					1000	1000		
	W 5000	2500					1000	1000		
# 0000004	R 6000	3800		1000			1000	1000		
	W 4000	2400					1000	1000		
# 0000005	R 7000	1400					1000			
	W 3000	600		1000 100			50	1000		
⋮	⋮	⋮		⋮	⋮		⋮	⋮	⋮	⋮

INFORMATION
コピー元論理ボリューム番号Aを論理ボリューム番号Bに
コピーすることを推奨します。ボリュームコピー実行
ウィザードの起動をしますか?
Yes(Y) No(N)

【図 13】

図 13



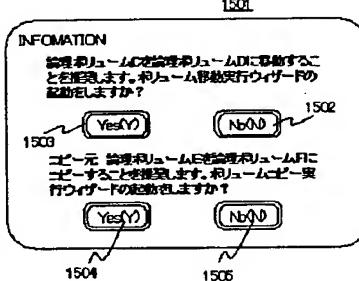
【図 14】

図 14

統計期間 : 1900-23/08/00 DAY		論理ボリューム 番号(LVOL#)	読み書き頻度 ($\times 10^3$)	同一単位ディスク 制御装置内チャネル バス頻度($\times 10^3$)	論理バス読み書き頻度 ($\times 10^3$)	#0	#1	#2	...	#N-2	#N-1	#N
# 000000	R 8000	2200		2200								
	W 2000	2000		2000								
# 000001	R 6000	0										
	W 6000	0										
# 000002	R 8000	0				8000						
	W 2000	0				2000						
⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
# N-2	R 0	0										
	W 0	0										
# N-1	R 4000	6200				600						
	W 4000	4000										
# N	R 7000	500				5500				500		
	W 3000	100				2000				100		

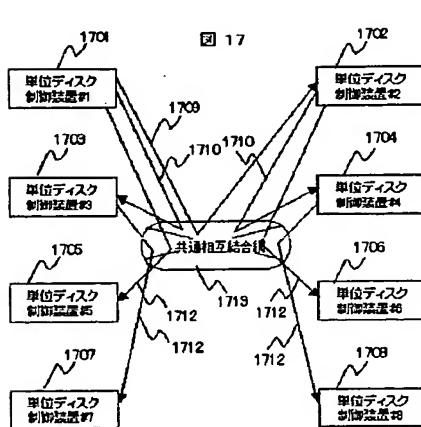
【図 15】

図 15



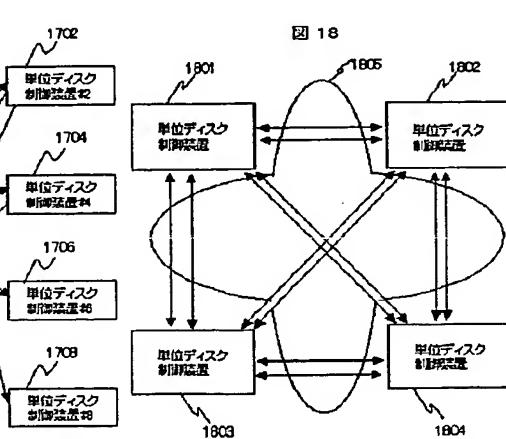
【図 17】

図 17



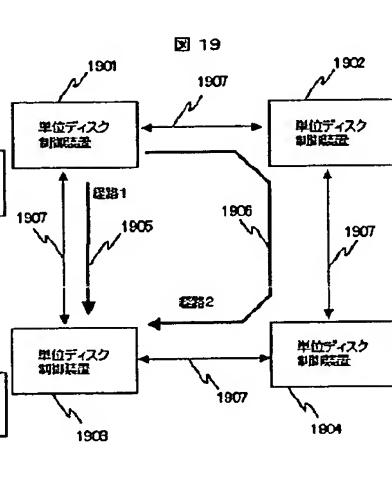
【図 18】

図 18

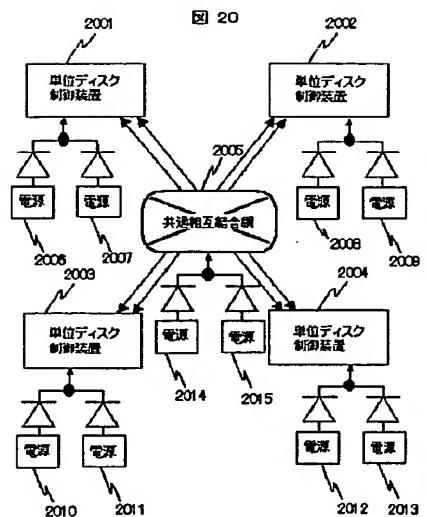


【図 19】

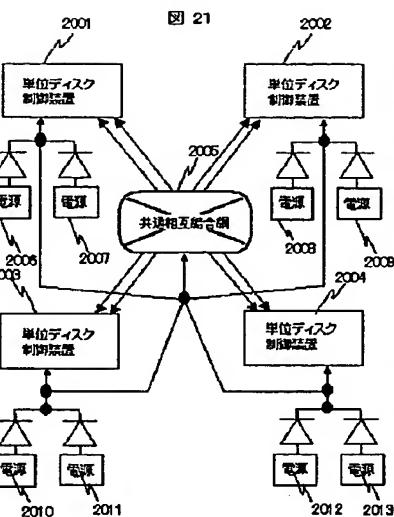
図 19



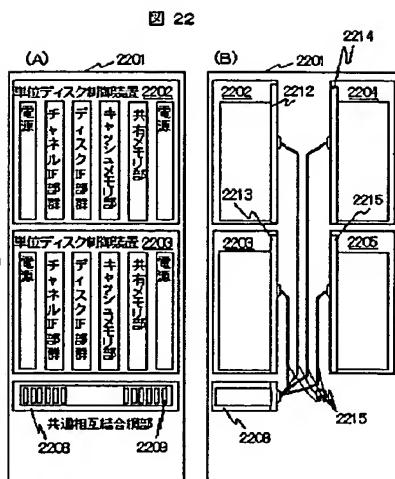
【図20】



【図21】



【図22】



フロントページの続き

(72)発明者 金井 宏樹

東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

Fターム(参考) 5B065 CA30 EK06